Chapter 10

BOATING BENEFIT ESTIMATION BASED ON CHOICES OF COMPLEMENTARY GOODS

This chapter outlines, in turn, the nature of econometric models of discrete (durable goods) choice and the procedures required to produce a monetary welfare measure from such models when one of the attributes of the choice (in this case water availability by category of boat) changes. After this overview the details of the Coast Guard boating data as supplemented are discussed, and a discrete choice model is estimated from the Coast Guard Survey Data. Finally, welfare measures associated with the effect of water quality improvements on the utility of discrete choice are derived.

THE DISCRETE CHOICE DEMAND MODEL

The consumer's qualitative choice problem in this context is whether to own a boat for pleasure cruising or hunting and fishing and, if the ownership decision is made, what kind of boat (power, sail, etc.) to own.

To avoid complicated combinatorial problems, assume that if the purchase decision is made, one and only one type of boat will be owned. Also, hypothesize that the available types of boats are qualitatively differentiated,[1] and that boat durables demand is influenced by the level of supply of the complementary public good, boatable water.

With these assumptions in mind, this section presents the analytics of the discrete choice demand model in a deterministic setting, a generalization of the theoretical exposition in Mäler 1974, also discussed by Small and Rosen 1981, and Hanemann 1981. Subsequently, the random utility (RUM) version of the model is outlined. The form of the distribution function (extreme value) of random taste variations is introduced to characterize the econometric RUM in terms of the conditional distribution of the observed outcomes, given an observed set of explanatory variables influencing the utility of each choice (McFadden 1974). Finally, methods for calculating welfare measures in the context of the budget-constrained RUM qualitative choice model are outlined, following Hanemann, 1982a, 1982b, 1983.

The Deterministic Choice Model

The discrete choice demand model can be written formally as a static utility maximization problem, subject to a budget constraint and constraints which impose mutual exclusivity and lumpiness in discrete goods consumption (Hanemann 1981):

$$\max.\ u(x_1,\ldots, x_N, \phi_1,\ldots, \phi_N, z) \qquad (1)$$

s.t.

$$\sum_i p_i x_i + p_z z = y$$

$$x_i x_j = 0, \text{ all } i \neq j$$

$$x_i = \bar{x}_i = 0 \text{ or } 1$$

$$z \geq 0$$

where x_i represents a durable good, p_i its capital service price, ϕ_i an index of the quality attributes of the i^{th} durable, and z a Hicksian composite commodity of undifferentiated quality with a normalized price p_z of unity.[2]

To get the model in (1) into a form amenable to econometric estimation it is convenient to derive the indirect utility function. First, the consumer's problem can be reformulated into a "surplus income" version of the utility function by substituting the budget constraint into the utility function in (1). The new problem equivalent to (1) is:

$$\max.\ u(x_1,\ldots,x_N,\ \phi_1,\ldots,\phi_N,\ (y - \sum_i p_i x_i)/p_z) \tag{2}$$

$$\text{s.t. } x_i x_j = 0, \text{ all } i \neq j$$

$$x_i = \bar{x}_i = 0 \text{ or } 1$$

$$z \geq 0$$

Now, if the consumer opts for purchase of durable good i his utility conditional on that decision, u_i recognizing the remaining constraints, is:[3]

$$u_i = u(0,\ldots,0,\ \bar{x}_i,\ 0,\ldots,0,\ \phi_i,\ 0,\ldots,0,\ (y-p_i\bar{x}_i)/p_z) \tag{3}$$

But, if the consumer decides not to own any durable, instead allocating all income, y, to the undifferentiated Hicksian composite commodity, utility conditional on that decision, u_z, is:

$$u_z = u\ (0,\ldots,0,\ 0,\ldots,0,\ y/p_z) \tag{4}$$

Both (3) and (4) above have prices and income (along with quality attributes) as arguments, and can be expressed as conditional indirect utility functions $v(\cdot)$, which Hanemann 1981 writes[4] as:

$$u_i = v_i(\phi_i,\ (y-p_i\bar{x}_i)/p_z)\quad i = 1,\ldots,N \tag{5}$$

$$u_z = v_z(y/p_z) \tag{6}$$

The functions are homogenous of degree zero in prices and income.

The consumer's decision can now be represented by a set of binary functions $\delta_1,\ldots,\delta_N,\ \delta_z$ which assume a value of 1 if the choice represented by the subscript (i=1 , ...N for discrete goods and z for exclusive consumption of the Hicksian good) is made and 0 otherwise. A particular choice will be made if the conditional utility derived from that choice dominates the conditional utility provided by any other choice. So, the binary variables are on-off switches related to the conditional indirect utility function by:

$$\delta_i(p,q,y) \quad \begin{array}{l} 1 \text{ if } v_i(\phi_i, (y-p_i\bar{x}_i)/p_z) \geq v_j(\cdot) \text{ and } v_z(\cdot), \text{ all } j \\ 0 \text{ otherwise} \end{array} \tag{7}$$

and

$$\delta_z(p,q,y) \quad \begin{array}{l} 1 \text{ if } v_z(y/p_z) \geq v_i(\cdot), \text{ all } i = 1,\ldots,N \\ 0 \text{ otherwise} \end{array} \tag{8}$$

This being the case, the ordinary unconditional demand functions from problem (2) are defined:

$$X_i(p, \phi,y) = \delta_i(p,\phi,y)\bar{x}_i = \delta_i(p,\phi,y) \tag{9}$$

$$z(p,\phi,y) = (y-\delta_i(p,\phi,y)\bar{x}_i p_i)/p_z \tag{10}$$

where, if any $\delta_i=1$ the quantity of z is fixed because of the budget constraint and if al1 $\delta_i=0$ the quantity of z is given as y/p_z.

Substitution of these demand functions into the utility function (1) yields the unconditional indirect utility function.[5]

$$v(p,q,y) = \max\,[v_1(\phi_1, (y-p_1\bar{x}_1)/p_z), \ldots, v_N(\phi_N, (y-p_N\bar{x}_N)/p_z), v_z(Y/p_z)] \tag{11}$$

For example, suppose the conditional indirect utility functions are linearly additive in attributes, service prices and income; and normalize p_z to unity:

$$V_i = \lambda(y-p_i\bar{x}_i) + \theta_i ;\ V_Z = \lambda y \qquad (12)$$

The parameter λ represents the marginal utility of expenditure, and the variable θ_i represents the way the quality characteristics of the i=1 ,..., N discrete commodities affect utility. In application, θ_i is often treated as a linear function of observable attributes, q. (See Hanemann 1981 for an extensive discussion). In this case:

$$\theta_i = \alpha_i + \beta_1 q_{1i} + \beta_2 q_{2i} + \ldots + \beta_m q_{mi} \qquad (13)$$

To solve this problem, the consumer must exhaustively evaluate all of the conditional indirect utility functions $v_i(\cdot)$ and $v_z(\cdot)$ to find the option which yields maximum utility, given his income.

Which option will he pick? If we rearrange the indirect utility functions $v_i(\cdot)$, (12) can be rewritten as:

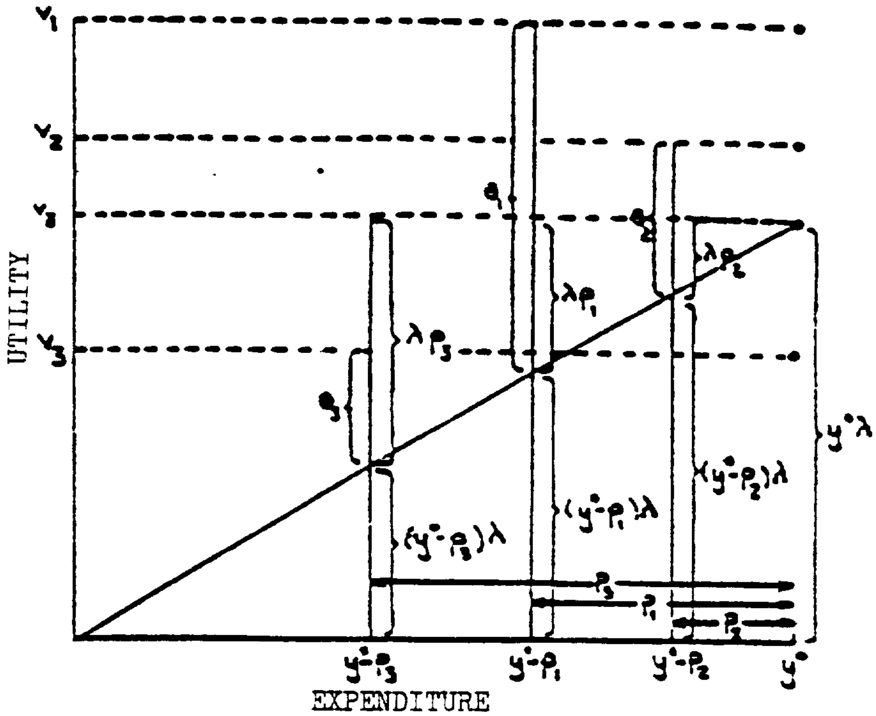
$$V_i = \lambda(y-\pi_i\bar{x}_i);\ V_Z = \lambda y \qquad (12')$$

where $\pi_i = p_i - (\theta_i/\lambda)$. Hanemann 1982 regards the π_i as “generalized” or quality corrected prices. Then the selection procedure can be arbitrarily regarded as a two step process. First, the consumer finds the dominant durable good among the i=1 , ...N durables, which will be the durable with the lowest value of π_i (ignoring ties). Then he compares the utility index of this choice, max $v_i(\cdot)$ to the utility index produced by exclusive consumption of the composite commodity $v_z(\cdot)$. He will choose to purchase the dominant durable and allocate the remainder of his income to the Hicksian composite if its value of π_i is negative and will consume only the Hicksian composite otherwise.

Several possibilities are displayed graphically in figure 10.1, where the individual’s income is y^*. The utility yielded by exclusive consumption of the Hicksian composite commodity is a linear function of income following (12) above. The marginal utility of expenditure, λ, is

Figure 10.1

A Framework for Analyzing Durable Goods Decisions



given by the ratio of the maximum utility attainable by spending all of his income on the composite, y_z^* to his income allotment, y^*. Three durable options are also shown, where $\pi_1 < \pi_2 < 0 < \pi_3$. Here, exclusive consumption of the Hicksian composite is preferred to diverting some income to the purchase of durable $\bar{x}_3$. But both $\bar{x}_1$ and $\bar{x}_2$ are preferred to exclusive consumption of the composite, and purchase of $\bar{x}_1$ is the dominant overall choice.[6]

<u>The Budget-Constrained Random Utility Yodel (RUM)</u>

The random utility model (McFadden 1974, 1981, 1982) arises by supposing that each individual has a utility function with a non-stochastic component reflecting representative tastes and an unobservable stochastic component. This latter component arises either from idiosyncratic interpersonal variations in individual tastes, intrapersonal stochastic choice behavior, or unobserved attributes of differentiated goods (McFadden 1974, 1981, 1982, Hanemann 1982).[7] The budget-constrained RUM adds the requirement that the measure of utility conditional on the j^{th} choice be interpreted as a conditional indirect utility function, as derived in the previous section (Eq. 11) (Hanemann 1981). In this formulation, explicit restrictions on the arguments and functional structure of the non-stochastic component of utility are implied.

Following McFadden (1974) and Amemiya (1981), let a typical individual's conditional utility corresponding to the j^{th} choice be a function of a vector of characteristics (Including prices) of alternative j facing individual i, z_{ij} (where from the previous development the vector Z_{ij} includes p_{ij}, p_Z and ϕ_{ij}) and characteristics w_i which vary across individuals but not across choices: These latter include the socioeconomic characteristics of the individual, like income, and characteristics of the environment like climate that may affect choice. That is, the average utility of choice j is:

$$\bar{v}_{ij} = \bar{v}(z_{ij}, w_i) = x_{ij}'\beta_j \tag{14}$$

where a linear form for the utility function is conveniently assumed, with x_{ij} a column vector of functions of z_{ij} and w_i, and β_j a parameter (column) vector common to the entire population. Each individual in the population thus has the random utility of choice j represented by the function:

$$u_{ij} = \bar{v}(z_{ij}, w_i) + \varepsilon_{ij} = x_{ij}'\beta_j + \varepsilon_{ij} \tag{15}$$

In (15) the stochastic errors ε_{ij} are independently and identically distributed with the type I extreme value distribution (Maddala 1983) with a cumulative distribution function $F(\varepsilon) = \exp(-e^{-\varepsilon})$. The errors represent the displacement of each individual's utility from the average as a function of the choice attributes and the individual's characteristics.

For example, denoting the choice of buying a boat as choice 1 and of not buying. as choice 0, the binary random utility comparison is:[8]

$$u_{i0} = x_{i0}'\beta_0 + \varepsilon_{i0}$$

versus (16)

$$u_{i1} = x_{i1}'\beta_1 + \varepsilon_{i1}$$

Then, individual i will own a boat if $u_{i1} > u_{i0}$ and will not own one otherwise.

Let (16) be written in a more detailed form consistent with the utility model of the previous section. Since the Hicksian choice involves an undifferentiated commodity, suppose there is one quality attribute influencing the utility of the boat choice, q1 and one general attribute, climate, influencing both choices, q_0:

$$u_{i0} = \lambda y_i + \beta_0 q_{0i} + \varepsilon_{i0} = \bar{v}_0 + \varepsilon_{i0}$$

versus

$$u_{i1} = \alpha_1 + \lambda y_i - \lambda p_{1i} + \beta_1 q_{0i} + \beta_2 q_{1i} + \varepsilon_{i1} \qquad (17)$$

$$= \bar{v}_{i1} + \varepsilon_{i1}$$

where, as before, λ represents the marginal utility of income, y.

Here the restrictive assumption is made that the individual's responses (the "marginal utilities") to some attributes (income, price) are invariant with respect to the choice, thus the nonsubscripted λ. However, the sensitivity of the choice utility to some environmental characteristics (climate) which vary over individuals but not choices is represented by the parameters attached to q_{0i} in the two equations, whereas the intercept value (0 or α_1) is alternative-specific, as is the influence of the quality attribute, q_{1i}.[9]

Now, defining the choice indicator I = 1 if the individual 13 observed to own a boat and I = 0 otherwise we have. from (17) the probability that an individual chooses alternative 1:

$$\Pr(I = 1) = \Pr(u_{i1} > u_{i0}) \tag{18}$$

$$= \Pr[(\varepsilon_{i0} - \varepsilon_{i1}) < (a, \; - \lambda p_{1_i} + (\beta_1 - \beta_0)q_{0_i} + \beta_2 q_{1_i})]$$

$$= \Pr[\varepsilon_{i0} < (\varepsilon_{i1} + \alpha_1 - \lambda p_{1_i} + (\beta_1 - \beta_0)q_{0_i} + \beta_2 q_{1_i})]$$

Assuming a type I extreme value distribution $F(\cdot)$ for the ε_{i0} and ε_{i1} it can be shown (Maddala 1983, p. 60) that, dropping the individual i subscript:

$$\Pr(I = 1) = \exp(\bar{v}_1)/(\exp(\bar{v}_1) + \exp(\bar{v}_0)) \tag{19}$$

$$= 1/(1 + \exp(v_0 - \bar{v}_1))$$

More generally for several choices rather than a binary choice, with I_j choice indicators which equal 1 if the j^{th} choice is made and zero otherwise:

$$\Pr(I_i = 1) = \exp(\bar{v}_i)/\sum_{j=1}^{J} \exp(\bar{v}_j) \tag{20}$$

$$= \frac{1}{\sum_{j=1}^{J} \exp(\bar{v}_i - \bar{v}_j)} = \frac{1}{1 + \sum_{j \neq 1}^{J} \exp(\bar{v}_j - \bar{v}_i)}$$

It is apparent from the form of (18), (19) and (20) that activity specific variables which vary in level across individuals but not choices are reflected in the estimated choice model only as parameter <u>differences</u>. And when a variable which varies across individuals but not choices is restricted to parameter equality across choices it drops out of the model altogether.

Particularly, income does not enter our "no income effects" version of the RUM model since the constancy of the marginal utility of expenditure, λ, is a maintained hypothesis (Hanemann 1982). Notably, prices and characteristics are the only variables influencing choice in the "no income effects" RUM model. Figure 10.2 shows, in a deterministic framework, how an increase in discrete goods prices from p_1, p_2 to p_1', p_2' makes the consumption of durables unattractive. Figure 10.3 contrasts a no-income effects model (Panels A and C) to a model where the income level affects choice probability because the marginal utility of income is not constant (Panels B and D). Panels A and C depict two individuals with incomes y_A and $y_B < y_A$ facing the same price set. Note that because the marginal

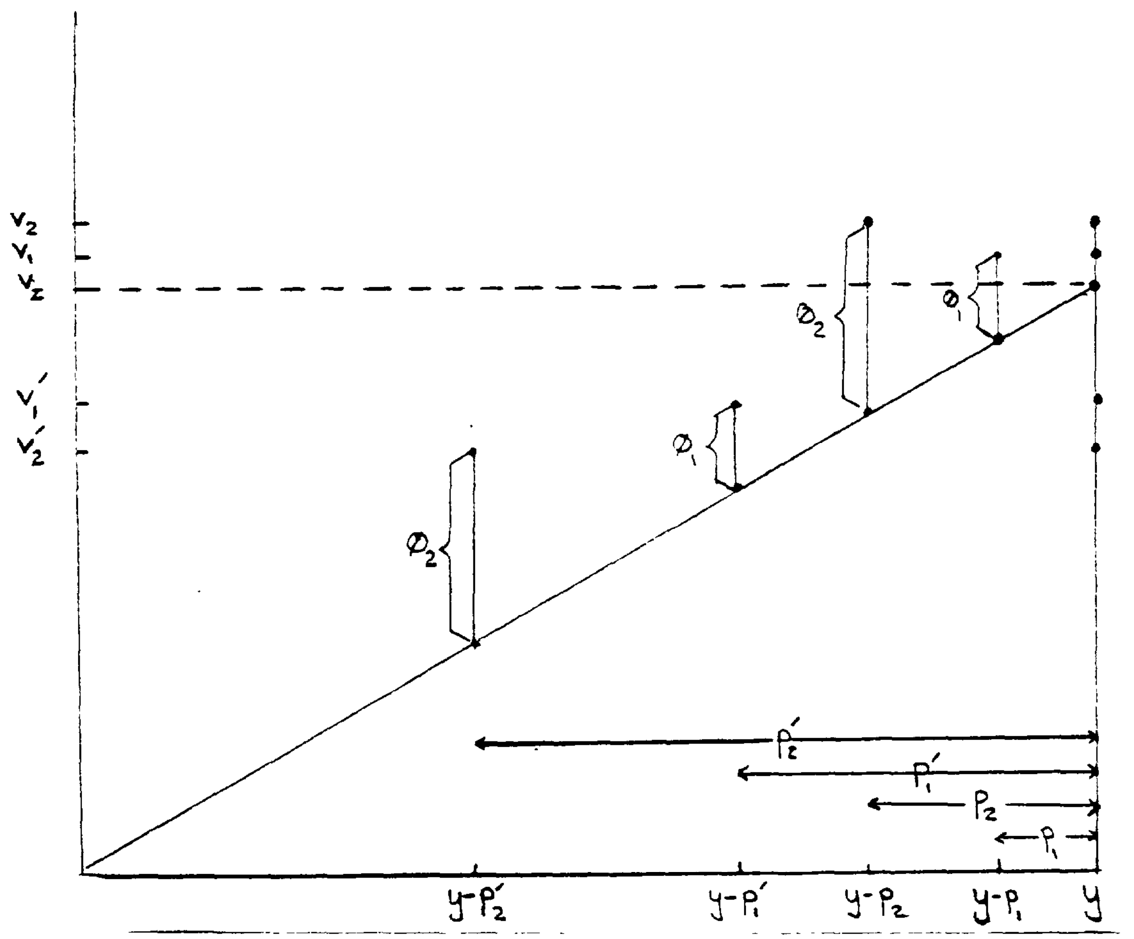
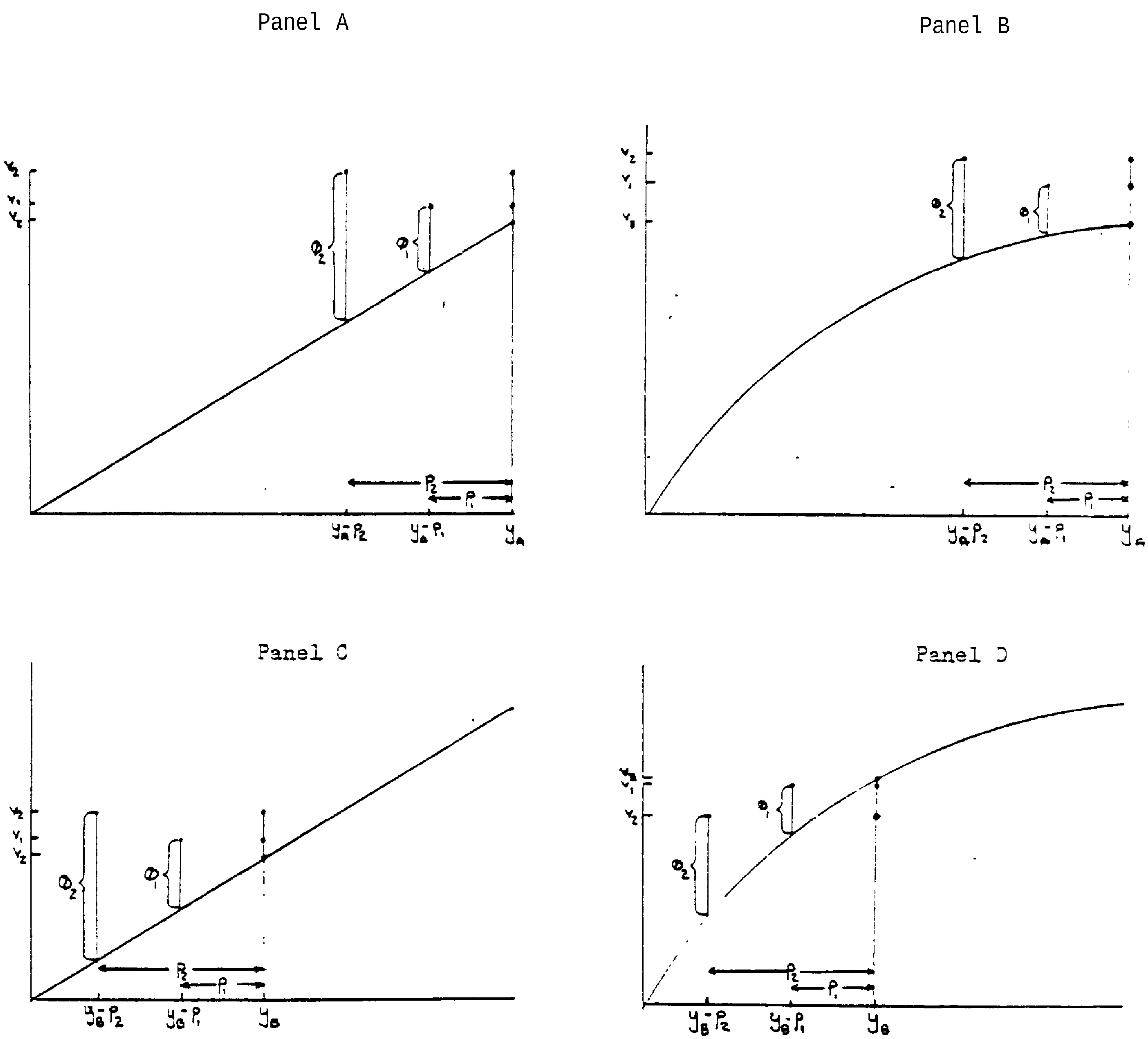


Figure 10.2
Changing Prices Affect Durable Goods Decisions

Figure 10.3

Contrasting Models With and Without Income Effects



utility of expenditure is constant, purchase of discrete good 2 is the preferred option at both income levels. But in Panels C and D the nonlinearity of the indirect utility function causes a purchase of good 2 to be dominant for an individual with income y_A, but exclusive consumption of the Hicksian composite to be the best choice for an individual with income y_B.

In our case the no income effects assumption is dictated by the lack of income data, as discussed below. The assumption is inconsistent with the possibilities that some durable purchases may be infeasible for households with sufficiently low incomes, or that marginal utilities of expenditure vary by income class (Deaton and Muellbauer, 1980, pp. 366-69). And, as a practical matter the inclusion of socioeconomic variables seems to produce more general models which are superior, by statistical criteria, to the more restrictive conditional logit node1 of choice outlined here (Hartman 1982, Henscher 1984). Rut in the absence of socioeconomic data, the advantage of the no income effects conditional logit RUM is ease in evaluation of the welfare effects of prize or attribute changes (McFadden 1981).

WELFARE ANALYSIS WITH THE BUDGET CONSTRAINED RUM

If a conditional logit model is employed to estimate the "no income effects" version of the budget constrained utility model, Hanemann (1981, 1982a, 1982b, 1983) has shown that monetary measures of welfare change due to changes in the level of non-price attributes can be obtained in a straightforward way. Hanemann’s exposition is an important extension and clarification of the welfare measures suggested by Small and Rosen (1981).

The critical assumption of the no income effects model is that the discrete goods are sufficiently unimportant that the income effects of a quality or price change are negligible, 30 the unconditional compensated demand function is adequately approximated by the unconditional ordinary demand function. This assumption also requires the conditional marginal utility of income, $\partial v_i/\partial y$ to be independent of p_i and θ_i (Small and Rosen, 1981, Hanemann 1982). When all is said and done this turns out to be the well known (and probably unrealistic) assumption of a constant marginal utility of expenditure (McKenzie 1983, pp. 87-88). This assumption requires all conditional indirect utility functions to have the form noted previously in Eq. 12, and the direct utility function must be additively separable in the Hicksian numeraire (McKenzie 1983, p. 88, Hanemann 1982, p. 8).

Under the maintained hypothesis of no income effects, the welfare effects of a change in the attribute or quality variables can be obtained via two routes, depending on whether the change is small and can be regarded as marginal or involves a large, non-marginal change.

Marginal Welfare Analysis

Define the compensating variation, CV, as the offsetting change in income necessary to make an individual as well off after a change in either prices or quality levels as he was before the change. Formally, in terms of the expenditure function, letting the 0 superscript represent the base situation and the 1 superscript the post-change situation:

$$CV = e(p^1,\phi^1,u^0) - e(p^0,\phi^0,u^0) \tag{21}$$

where $e(\cdot)$ represents the minimum expenditure required to reach the stated utility level, given the quality and prize vector and $e(p^0,\phi^0,u^0) = y$, the individual's income. Compensating variation can also be defined in terms of the indirect utility function, $v(\cdot)$ which is the inverse of the expenditure function: In general:

$$v(p^1,\phi^1,y-CV) = v(p^0,\phi^0,y) = u^0 \tag{22}$$

or

$$v(p^1,\phi^1,y-CV) - u^0 = 0 \tag{23}$$

Suppose only quality changes. Holding prices constant and implicitly differentiating (23)

$$-\frac{\partial v(\cdot)}{\partial CV}\,dCV + \frac{\partial v(\cdot)}{\partial \phi}\,d\phi = 0 \tag{24}$$

or

$$\frac{dCV}{d\phi} = \frac{\partial v(\cdot)/\partial \phi}{\partial v(\cdot)/\partial CV} = \frac{\partial v(\cdot)/\partial \phi}{\partial v(\cdot)/\partial y} \tag{25}$$

For simplicity suppose as before that the ϕ_i's are scalar quality indices, each associated with its own conditional indirect utility function $\bar{v}_i(\cdot)$. Then applying the chain rule to the general formulation in (25) for a change in ϕ_i:

$$\frac{dCV}{d\phi_i} = \frac{(\partial v(\cdot)/\partial \bar{v}_i)(\partial \bar{v}_i/\partial \phi_i)}{(\partial v(\cdot)/\partial \bar{v}_1)(\partial \bar{v}_1/\partial y) + \cdots + (\partial v(\cdot)/\partial \bar{v}_N)(\partial \bar{v}_N/\partial y)} \tag{26}$$

$$= \frac{(\partial v(\cdot)/\partial \bar{v}_i)(\partial \bar{v}_i/\partial \phi_i)}{\sum_{i=1}^{N}(\partial v(\cdot)/\partial \bar{v}_i)(\partial \bar{v}_i/\partial y)}$$

Hanemann 1983 employs the result that in the conditional logit model the moments of the distribution function of the error term. are independent of p, ϕ, and y, so the probability of choice i, P_i, equals $\partial v(\cdot)/\partial \bar{v}_i$. Then (26) is:

$$\frac{dCV}{d\phi_i} = \frac{P_i(\partial \bar{v}_i/\partial \phi_i)}{\sum_{i=1}^{N} P_i\,\partial \bar{v}_i/\partial y} \tag{27}$$

Since the "no income" effects model has $\partial\bar{v}_i/\partial y = \partial\bar{v}_j/\partial y$ for all i, j, and by definition the sum of the P_i equals 1, we get:

$$\frac{dCV}{d\phi_i} = P_i \frac{\partial\bar{v}_i/\partial\phi_i}{\lambda} \tag{28}$$

where λ is the negative of the parameter estimate attached to the generic durable good service price variable in the logit model and $(\partial\bar{v}_i/\partial\phi_i)/\lambda$ is the marginal rate of substitution between quality and money, MRS_i, conditional on the choice of alternative i. (A similar result is given by Hensher and Johnson, 1981 , pp: 243-245).

So, up to a first-order approximation the marginal measure of the benefit of a change in ϕ_i is:

$$CV_i = \left. dCV/d\phi_i \right|_{\phi_i=\phi_i^o} \Delta\phi_i \tag{29}$$

Yore generally, if all qualities change simultaneously:

$$CV = \sum_{i=1}^{N} CV_i = \sum_{i=1}^{N} P_i MRS_i \Delta\phi_i \tag{30}$$

where the summation is over N durable choices where the Hicksian composite is unaffected by quality changes.

The formula in (30) yields a quick first cut approximation to the benefits of one or more quality changes, and is particularly easy to calculate since all it requires are estimates of the initial selection probabilities, the parameter estimates, and the hypothesized quality changes. However, a more accurate welfare measure can be produced which does not rely on a first order approximation.

Non-Marginal Welfare Analysis

In the no-income-effects RUM model the demand function for a durable good was given above in Eq. 9 as $x_i = \delta_i \bar{x}_i$ where δ_i was the utility maximization discrete choice index: In the RUM model the expected value of the discrete choice index is the probability P_i, that choice i will be made. The probability P_i can also be interpreted as a fractional consumption rate of the durable:

$$E(x_i) = P_i \bar{x}_i, \text{ where } \bar{x}_i = 1 \tag{31}$$

Now, we also know that P_i equals $\partial v(\cdot)/\partial \bar{v}_i$ so again using the chain rule and following Hanemann 1981:

$$\frac{\partial v(\cdot)}{\partial p_i} = \frac{\partial v(\cdot)}{\partial \bar{v}_i} \cdot \frac{\partial \bar{v}_i}{\partial p_i} = P_i \frac{\partial \bar{v}_i}{\partial p_i} \tag{32}$$

Similarly, differentiating the unconditional indirect utility function with respect to income, y,

$$\frac{\partial v(\cdot)}{\partial y} = \sum_{i=1}^{N} (\partial v(\cdot)/\partial \bar{v}_i)(\partial \bar{v}_i/\partial y) \tag{33}$$

which, from the no income effects assumption of $\partial v_i/\partial y = \partial v_j/\partial y$ can be written more simply as

$$\frac{\partial v(\cdot)}{\partial y} = \sum_{i=1}^{N} P_i \partial \bar{v}_i/\partial y = \partial \bar{v}_i/\partial y \tag{34}$$

Then, from Roy's identity we know that the demand equation for any x_i is given by the negative of the ratio of the derivative of the indirect utility function with respect to p_i to the derivative with respect to y, or that:

$$p_i \bar{x}_i = - \frac{p_i \, \partial \bar{v}_i / \partial p_i}{\partial v_i / \partial y} \tag{34.a}$$

and

$$\bar{x}_i = - \frac{\partial \bar{v}_i / \partial p_i}{\partial v_i / \partial y} = 1 \tag{34.b}$$

so $\quad \partial \bar{v}_i / \partial y = - \partial \bar{v}_i / \partial p_i = \lambda \qquad (34.c)$

Importantly, (34.c) tells us that with a linear specification of the choice index in our logit model, the negative of the estimated parameter on the generic cost variable is an estimate of the marginal utility of expenditure (Small and Rosen, 1981, p. 126). (This a roundabout way of proving what is obvious by inspection of figure 10.1 and Eq. 12 above.) So, if some measure of the change in the utility index due to a change in any of its arguments can be found, that change is easily converted into a monetary measure using the model's parameter estimate for λ.

In the conditional logit model it so happens that the expected value of the consumer's maximum utility level (i.e., the expected value of his indirect utility function) can be recovered from the estimated model. It is a well known result (Ben-Akiva and Lerman 1979, p. 663, McFadden 1982, p. 11, McFadden 1981, p. 222) that the expected value of the maximum indirect utility in the logit model is:

$$\begin{aligned} E\{\max v(\cdot)\} &= E\{\max[v_1+\varepsilon_1,\ldots,v_N+\varepsilon_N,v_z+\varepsilon_z]\} \\ &= E\{\max[\lambda y-\lambda p_1+\phi_1+\varepsilon_1,\ldots,\lambda y-\lambda p_N+\phi_N+\varepsilon_N,\lambda y+\varepsilon_z]\} \\ &= \lambda y+E\{\max[-\lambda p_1+\phi_1+\varepsilon_1,\ldots,-\lambda p_N+\phi_N+\varepsilon_N,\varepsilon_z]\} \\ &= \lambda y+\ln(\sum_i e^{w_i}+1)+0.5772 \end{aligned} \tag{35}$$

where p_z is normalized to unity, and the subscript for individuals is dropped. $E\{\cdot\}$ denotes the expectation operator, $w_i = -\lambda p_i + \phi_i$, $w_Z = 0$ so $e^0 = 1$, and 0.5772 is the mean of the standardized Type I Extreme Value Error Distribution (Johnson and Kotz 1970).

The expression $\ln(\sum_i e^{w_i}+1)$ is the natural logarithm of the denominator of the conditional logit model given in Eq. 20 above. Then, with compensating variation defined in (22), the utility index defined in (35) measured at the base (w_i^o) and post policy (w_i^1) levels of price and/or quality, arguments can be employed to produce the welfare measure (Hanemann 1982):

$$v(p^1,\phi^1,y+CV) = \lambda(y+CV) + \ln(\sum_i e^{w_i^1}+1) + 0.5772$$

$$= v(p^o,\phi^o,y) = \lambda y + \ln_i(\sum e^{w_i^o}+1) + 0.5772 \qquad (36)$$

so

$$CV = (\ln(\sum_i e^{w_i^o}+1) - \ln(\sum_i e^{w_i^1}+1))/\lambda.$$

THE DATA FOR THE DISCRETE CHOICE MODEL

The Nationwide Boating surveys (U.S.D.O.T. 1978) were conducted by telephone in 1973 and 1976 by the U.S. Coast Guard in an effort to gather detailed information on boat ownership and boating accidents. Only the 1976 survey is available on tape, and is the data set used herein.

Via a two-stage random cluster sampling Plan an attempt was made to produce a nationally representative probability sample.[10] The survey procedure involved screening questions to discern between households which owned and/or operated a boat in 1976 and those which did not. While an extensive set of follow-up questions was asked of owner/operator households identified in the screen, no further information was collected from households which neither owned nor contained members who operated a boat in 1976.

The screening questions identified the number of boats owned by a household, the number of persons in the household participating in boating, the number of boat rentals by the household over the year, and the number of household members who operated a boat in 1976. (Participating in boating was defined as operating, riding in or waterskiing behind a boat.) Of the series of follow-up questions, most were concerned with the characteristics of the boat(s) owned, safety considerations, and accident experience of owner-operator households. The only socioeconomic data collected were the age and occupation of the household's primary boat operator. Notably, the survey did not elicit socioeconomic data on households without boat operators, and no household income information was collected on any of the households contacted.

because one of the primary purposes of the survey was to produce information on boating accidents and total hours of boating exposure to the possibility of accidents, the only measure of boating intensity sought was the amount of tine a boat was operated over the year. No information was gathered on the amount of time (recreation days) that any individual boater or operator participated in boating recreation, and such an individual-specific measure cannot be inferred from any of the boat use information in the survey. However, it is possible to construct a measure of annual boating trips per household, which is employed in the next chapter.

Finally, the survey did not determine where boating activities took place, nor did it distinguish between marine and freshwater boating. Nonethe1ess, the survey can be used to create dependent variables for estimation of a conditional logit model of individual durable goods choice.

Dependent Variables

To use the Coast Guard Survey data in a conditional logit analysis of boat durable ownership choices, we identified six categories of durable goods choice: The first category, non-ownership, includes households which did not own a boat. The remaining five mutually exclusive categories refer to boat ownership distinguished by the degree of water contact (small versus large boats) and the means of boat propulsion (motor, sail, and paddle or oar). For estimation a subsample of observations was drawn from the full sample to economize on computational cost. Only single boat owning households or two-boat households whose primary boat was a large craft were included in our samples. The number and frequency of observations by choice category are shown in table 10.1. These mutually exclusive categories were used to create individual observations in the Coast Guard Survey in binary dependent variable form, the value 1 being assigned when an observation belongs to a category and zero otherwise.[11]

As is common in transport model choice analysis, this dependent variable data must be supplemented by independent variables characterizing the costs and attributes of each of the choices in the choice set. Because this information is not available on an individual-specific level, the construction of state averages for the various capital service prices and state (or county) averages for the attributes (availability of boatable-quality water) is described below. This procedure is legitimate when individual observations are used as dependent variables and all independent variables are measured as zonal averages, because consistent parameter estimates can be obtained, assuming that the independent variables are symmetrically distributed within each zone (McFadden and Reid 1975).

Table 10.1. Observations by Boat Ownership Category from the Coast Guard Survey

Category	Number of Available Observations: Full Sample	Number of Available Observations: Sub-Sample	Percent of Total Available: Full Sample	Percent of Total Available: Sub-Sample
Non-Ownership	20624	6162	88.76	87.78
Large Power Boat Ownership	677	236	2.91	3.36
Large Sail Boat Ownership	52	23	0.22	0.33
Small Power Boat Ownership	1581	485	6.81	6.91
Small Sail Boat Ownership	82	31	0.35	0.44
Small Other (Canoes, etc.)	221	83	0.95	1.18
Total Available	23237[a]	7020[b]	100.00	100.00

Notes:

a. Net of 449 multiple boat owning households.

b. From the available sample of 23237 observations, 4750 were not usable due to missing pollution data (residents of Alabama, Florida, Georgia, Indiana, Maryland, Massachusetts, Nevada, New York, and West Virginia). The usable sub-sample of size 7020 represents a systematic draw from the usable sample of 18487 observations.

Independent Variables.

While inclusion of the capital service price (see Appendix A) as a generic variable in the choice model is dictated by economic theory, no such clear guidelines are available for the specification of the relevant quality or attribute variables belonging in the model. However, it is reasonable to suppose that the availability of fresh, marine and Great Lakes water somehow matters for all choices involving boat ownership. The problem is how to measure availability, and how to incorporate pollution in the model.

As described in chapter 2 it is reasonable to use a physical measure of gross county or state freshwater availability (U.S. Department of Commerce 1981) as a proxy for the impedance to boat ownership arising from the owner's need to travel to freshwater locations where boating can be enjoyed.[12] However, an analogous area of water per unit land area measure of marine or Great Lakes availability cannot be similarly defined because of the vast expanse of these water bodies. Instead, the distance from each respondent's home, county centroid to the centroid of the closest marine or Great Lakes county can be computed from location information in U.S. Department of Commerce 1979.

These gross availability measures do not reflect limitations on boating due to pollution and other reasons. Particularly, boating activity may be restricted by shallow water, weeds, and commercial sea lanes, not to mention the lack of availability of boat ramps and marinas. Second, pollution may (or may not) discourage boat ownership. Neither of these sorts of limitations is easy to quantify, although an attempt to do so is reported in Dyson's 1984 survey of state officials included as appendix C to chapter 5 above. However a close look at these figures suggests they may not be very reliable.

Restrictions on availability for reasons <u>other than</u> pollution effectively increase the requisite travel distance. So, our gross freshwater availability figures can be adjusted downward by the percentage of water unavailable for boating for reasons other than pollution reported by Dyson.[13] Similarly, the gross distance to the nearest marine coast for each individual can be inflated by an area-weighted average of the non-pollution limitations for bays, estuaries, and coastal waters reported by Dyson.[14] The Great Lakes distances were not adjusted, because in our judgement the non-pollution limitation figures in this category given by Dyson appear implausible and mutually inconsistent.[15] While there is the possibility that the Dyson data on the percent of water unavailable for reasons other than pollution is seriously flawed, it seems preferable to adjust the "gross" freshwater availability and marine distance variables rather then employ them directly without adjustment.

The limitations on availability for pollution-related reasons are also problematical, first for what they truly measure and second, how they should be reflected in the model. The pollution data gathered in the survey reflect survey respondents subjective impressions of the amount of total water in their states which is unsuitable for boating because of pollution. Boaters, may not make the same evaluations or reflect such evaluations in their behavior. Indeed, they may not even be deterred in any way from owning a boat due to patches of "unsuitable" water, however it is measured. In this case, the method here being explored would produce zero benefits of pollution control.

Some objective characterization of water by class, or quality index, reflecting observable attributes such as biochemical oxygen demand, presence of oil slicks, floating debris, algae, odor etc. would be preferable to these subjective data. Unfortunately, there exists no data base with complete national coverage of marine and freshwater that would permit the construction of such a measure. Thus, Dyson's data is the best (only) available, and all models reported below are estimated with the Dyson pollution and non-po11ution limitation measures. Apart from pollution, there is at least one environmental variable which one might expect a priori to be particularly relevant to the boat ownership decision, the harshness of the weather. The longer the expected boating season, the more attractive the boat ownership option.

Finally, models employing the same structural specification are estimated at two levels of spatial aggregation for the freshwater availability variable - the state and the county - to demonstrate model sensitivity to independent variable measurement. The freshwater availability variable itself is alternatively measured as either square miles (acres) of water per square mile (acre) of total surface area or in distance-proxy form as the square root of the reciprocal the former. While the models are non-nested, the county specification with either freshwater availability measure is preferable on a priori grounds.

Table 10.2 defines the cost, availability, pollution and other environmental variables deemed relevant to choice. The last column of table 10.2 indicates the structure of the estimated models which is discussed in detail below. Table 10.3 provides sample means and standard deviations for the independent variables.

Table 10.2. Independent Variables

Variable Name	Description	Source[a]	Hypothesized Effect on Probability of Boat Ownership[a]	Model
HDD	Annual heating degree days, an index of the extent of negative departure of the average temperature from a base of 55°F. An activity-specific variable in the no ownership choice category.	DOC 82	(-)[b]	1,2,3
COST	Rental price in 1976 dollars of boat capital services normalized by a cross-sectional index of the composite commodity price. A generic variable.	Appendix A	(-)	1,2,3
DIST	Distance in miles Prom each individual's county of residence to the closest marine or Great Lakes coast.	Calculated as Euclidian distance based on origin and destination coordinates.	(n.a.)	n.a.
ACRE(S) or ACRE(C)	Square miles (acres) of fresh water per square mile (acre) land area measured at either the state (S) county (C) level multiplied by 10 for scaling. Indexed to individuals by state of residence.	DOC 81	(n.a.)	n.a.
LIMITM	Fraction of marine and Great Lakes water area unavailable for reasons other than pollution, indexed to individual's destination and boat choice category. Uses area-weighted average of the bays, estuaries and coastal figures reported in Dyson for marine waters.	Dyson	(n.a.)	n.a.

Table 10.2 (continued)

Variable Name	Description	Source[a]	Hypothesized Effect on Probability of Boat Ownership[a]	Model
LIMITF	Fraction of freshwater area unavailable for reasons other than pollution, indexed to individuals by state of residence.	Dyson	(n.a.)	n.a.
ADIST	Limitation-adjusted distance to nearest available marine or Great Lakes coast. Equal to DIST divided by (1-LIMITM).	n.a.	(-)	1,2,3
AACRE(S) or AACRE(C)	Limitation-adjusted freshwater acreage per unit land area.	n.a.	+	1,2,3
POLMG	Fraction of total marine or Great Lakes water area which is unsuitable for boating due to pollution, indexed to individual's destination and boat choice category. Area-weighted average of bays, estuaries and coastal pollution reported in Dyson used for marine waters.	Dyson	(n.a.)	n.a.
POLFR	Fraction of total fresh-water area which is polluted, indexed to individual's state of residence and boat choice category.	Dyson	(n.a.)	n.a.
POLDIST	Increment to adjusted distance to marine or Great Lakes coast due to pollution reasons. Equal to ADIST multi-plied by POLMG.	n.a.	(-)	1,2

Table 10.2 (continued)

Variable Name	Description	Source[a]	Hypothesized Effect on Probability of Boat Ownership[a]	Model
POLACRE(S) or POLACRE(C)	Decrement to adjusted freshwater acres per unit land area due to pollution reasons. Equal to ACRE multiplied by POLFR.	n.a.	(-)	1,2
DLP DLS DSP DSS DC	Activity-specific dummy (D) variables for large power (LP), large sail (LS), small power (SP), small sail (SS), or canoe kayak (C) choices.	n.a.	?	1,2,3
POISSON(S)	$AACRE(S)^{-1/2}$, a state freshwater distance proxy	n.a.	(-)	1,2,3
POISSON(C)	$AACRE(C)^{-1/2}$, a county freshwater distance proxy	n.a.	(-)	1,2,3
POLPOIS(S)	Increment to state freshwater distance proxy due to pollution, Equal to POISSON(S) multiplied by POLFR.	n.a.	(-)	1,2
POLPOIS(C)	Increment to county freshwater distance proxy due to pollution, Equal to POISSON(C) multiplied by POLFR.	n.a.	(-)	1,2

Notes:

a. The designation n.a. indicates construction of a final variable based on the intermediate variables whose sources are indicated.

b. In estimation this variable appears as an activity-specific continuous variable in the non-ownership choice category for data-formatting convenience. The sign of the estimated parameter should therefore be positive, indicating that residence in colder climates increase the probability of non-ownership.

c. Model are either the full model (1), the environmental model (2) or the skeptical model (3) discussed in the text. The designation n.a. indicates an intermediate variable employed in the construction of variables ultimately used in estimation under Model column.

Sources:

DOC 81: U.S. Department of Commerce, Bureau of the Census. 1981. 1980 State/County Area Measurement, (unpublished data).

DOC 82: U.S. Department of Commerce, Bureau of the Census. 1982. Statistical Abstract of the United States. 103rd. ed., Washington, D.C., GPO.

Dyson: Pamela J. Dyson. 1984. Recreational Water Availability in the United States, Washington, D.C.: Resources for the Future. Appendix 5.C in this report.

Table 10.3. Variable Statistics

Variable[a]	Sample Mean	Standard Deviation
HDD	5356.75	2176.07
COST, Large Power	3373.08	262.59
COST, Large Sail	7704.40	601.07
COST, Small Power	661.29	51.46
COST, Small Sail	539.84	41.83
COST, Canoes	70.83	5.54
ADIST, Large Boats	183.73	206.01
ADIST, Small Boats	187.41	215.05
AACRE(S), Large Boats	1943.37	1543.50
AACRE(S), Small Boats	2097.78	1497.21
AACRE(C), Large Boats	2121.05	2917.03
AACRE(C), Small Boats	2260.30	2953.52
POLDIST, Large Boats	0.12	0.27
POLDIST, Small Boats	2.82	7.74
POLACRE(S), Large Boats	59.21	150.56
POLACRE(S), Small Boats	83.76	154.16
POLACRE(C), Large Boats	57.21	206.93
POLACRE(C), Small Boats	83.23	227.81
POISSON(S), Large Boats	8.81	3.53
POISSON(S), Small Boats	8.11	2.72
POISSON(C), Large Boats	15.81	33.78
POISSON(C), Small Boats	14.57	31.83
POLPOLS(S), Large Boats	0.20	0.32
POLPOIS(S), Small Boats	0.35	0.60
POLPOIS(C), Large Boats	0.30	0.65
POLPOIS(C), Small Boats	0.61	2.02

Note:

a. An S in parenthesis designates the state level of variable measurement, and a C the county level.

Model Specification: The Role of Pollution and Other Issues

There are at least two possible hypotheses on the influence of freshwater and marine pollution on the desirability of boat ownership: The environmentalist position regards the existence of pollution as effectively withdrawing water from the usable for boating category, so pollution reductions are hypothesized to have the same positive effect on the demand for the boat durable that the creation of new impoundments would have. At the opposite extreme, a skeptical position would maintain that pollution has no effect whatsoever on the desirability of boat ownership, other things equal.

A more balanced approach to the question would admit that pollution may or may not affect the utility of some or all boat purchase options. This neutral view treats the role of pollution as a statistically testable hypothesis by forming a more general conditional logit choice model specification than either the environmentalist or skeptical views would admit. Unfortunately, while the validity of either of the narrow positions vis-a-vis the general model can be tested statistically, the possibility exists that the restrictions of the null hypotheses of both of the narrow models may not be rejected in separate tests against the full model. The conundrum raised by the possibility of two plausible but non-nested narrow models is in general irreconcilable.

While the narrow model with the highest likelihood function value can be taken to represent the preferred specification, (Amemiya 1981), this model discrimination criterion (variously labelled the Sargan test or Akaike’s Information Criterion) is not really a statistical test with known properties. Rather, it should be successful "on average" presuming one of the models in the comparison set is indeed the true model.

To demonstrate how this situation arises, assume a linear specification of the utility index v_i and represent freshwater availability as Q, distance to the nearest coastline as D, the fraction of water area polluted as P_Q, the fraction of marine water polluted as P_M, and let all other influences on the choice index (cost, etc.) be collapsed for simplicity into an augmented intercept, K. Then the choice index in the general model is

$$v_i = K_i + \beta_1(Q_i) + \beta_2(Q_i \cdot P_i) + \beta_3(D_i) + \beta_4(D_i \cdot P_i) \qquad (37)$$

where $\beta_1 > 0$, $\beta_2, \beta_3, \beta_4 < 0$. In this representation pollution increases the expected distance of travel to the recreation destination or, otherwise said, the travel-associated cost of boat ownership and use.[16] The environmentalist model hypothesizes that $\beta_1 = -\beta_2$ and $\beta_3 = \beta_4$ while the skeptical model restricts $\beta_2 = \beta_4 = 0$. Obviously, then, the environmentalist and skeptical models are non-nested, and the former in effect "guarantees" a benefit from pollution reductions if the parameters of the reduced model are significant and properly signed.[17] Finally, in addition to the way freshwater availability is measured (acres per acre of land or the negative square rooting thereof) and the way pollution enters the model, two other specification issues remain.

The first is whether the parameters attached to freshwater availability, distance, and the pollution variables are constrained to be the same across boat ownership choices or are allowed to vary according to, say, boat size (large versus small). The former model is a restricted version of the latter, and we estimate both below, where models allowing parameter variation across choice categories are referred to as the <u>General</u>

models and those hypothesizing parameter equality are referred to as Restricted models. Within either class, environmentalist, skeptic or full models defining the role of pollution can be estimated.

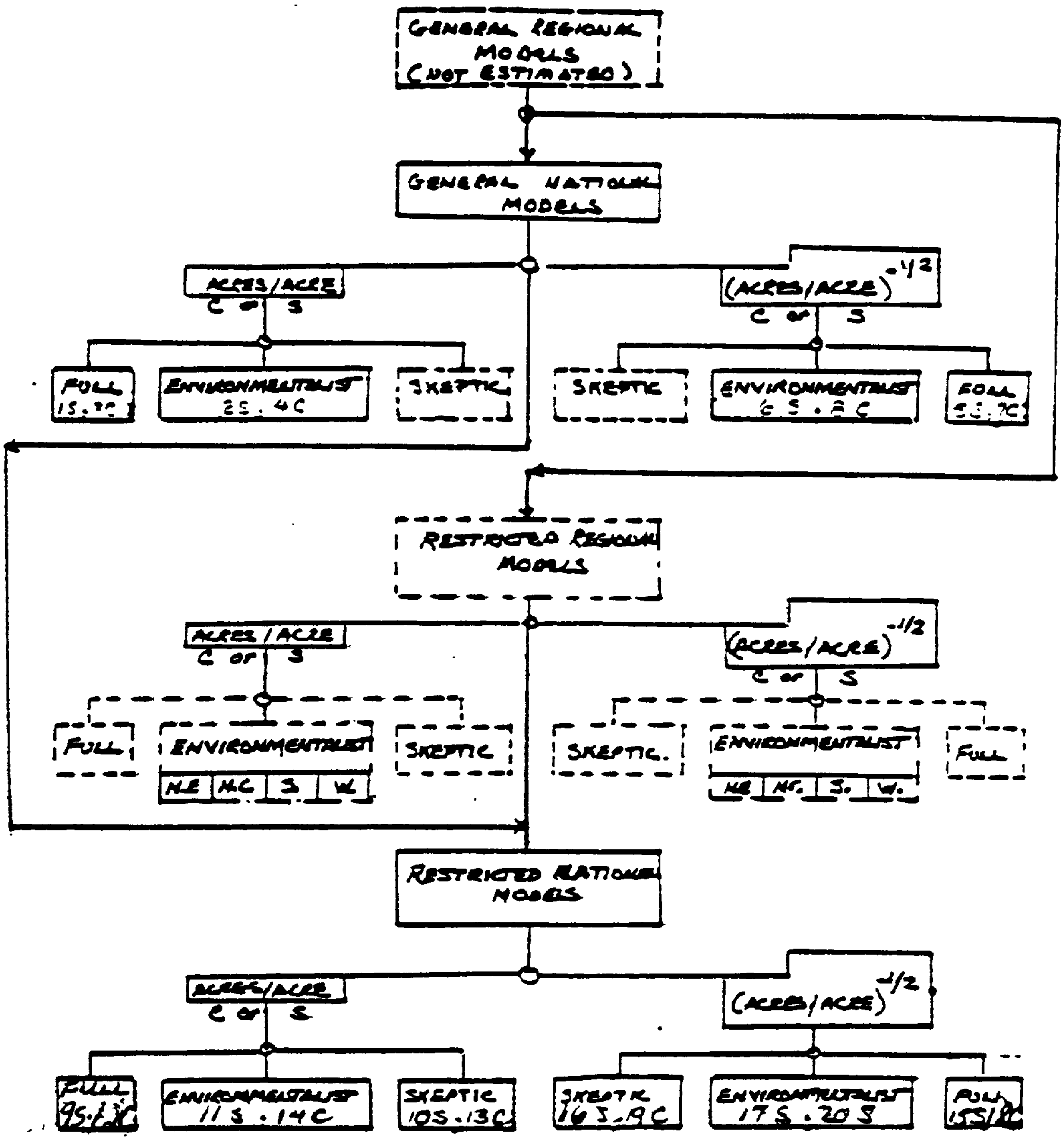
The second issue is whether or not a nationally representative "average" indirect utility function can legitimately be hypothesized, especially when there is reason to believe that preferences may well be regionally differentiated in some systematic way which cannot be captured by continuous variables measuring the socioeconomic attributes of individuals. When this additional complication is introduced on top of the model specification issues previously discussed an obvious combinatorial problem exists, exploding the number of potentially estimable models, But, because our principal goal is national benefit estimation, attention is confined to models which maintain a nationally shared representative or average indirect utility function.

The accompanying flow chart of figure 10.4 outlines the alternative model specifications and the legitimate (nested) hypothesis testing paths. As mentioned above, for purposes of economy some specifications, especially those unlikely to provide positive benefits, were not estimated. Models are numbered to key to the tables that follow where an (S) following a model number indicates that the state level of measurement was used for freshwater availability and a (C) indicates county measures.

EMPIRICAL RESULTS: BOAT DURABLE CHOICE MODELS

Based on a systematic subsample of size 7020 drawn from the full sample of usable observations, various six-choice conditional logit models can be estimated to explain the allocation of observations by choice category observed in table 10.1 above.

Figure 10.4. Choice Model Specifications



C, S = LEVEL OF MEASUREMENT OF FRESHWATER AVAILABILITY. C = COUNTY, S = STATE

⟶ VALID HYPOTHESIS TEST PATH

- - - MODEL NOT ESTIMATED.

The estimated parameters of the alternative choice models appear in tables 10.4 through 10.7. Full and Environmentalist specifications of the National General model which permit parameter inequality for availability, distance and pollution variables across large and small boat choice categories appears in table 10.4. There, freshwater availability is measured at the state or county level as acres per surface acre. Table 10.5 presents the same General model specifications, but measures availability as the negative square root of acres per acre at the state or county level of measurement following the argument in Vaughan and Russell 1984. Tables 10.6 and 10.7 present Full, Skeptic and Environmentalist specifications of the Restricted National models which force parameter equality across large and small boat categories in response to distance to marine or Great Lakes coastline, freshwater availability, and pollution. Table 10.6 uses state or county freshwater acres per surface acre to measure availability, and table 10.7 uses the negative square root transform.

As these tables are extensive a discussion of specific parameter estimates for each model would be unreasonably tedious.[18] However some general patterns are obvious in the national models. First every model produces significant negative parameter estimates consistent with prior expectations for the capital service price variable. Second, the distance to marine or Great Lakes coastline variable is never significant at conventionally accepted levels in any of the models, while the freshwater availability measures often (but not always - see models 18, 19 and 20) are. Moreover, only two of the eight Full models (3,7) produce significant coefficients on any pollution variables, and these on the freshwater acreage are not available for large boating due to pollution when county (not state) availability data is used. This result, given the overwhelming

Table 10.4. General National Choice Models: Acres Freshwater for Acre Surface Area Availability Proxy (absolute value of "t* statistic) in parenthesis

Variable	State Level Freshwater Acres		County Level Freshwater Acres	
(S)=State, (C)=County Large=Big Boat Small=Small Boat	Full 1	Environmentalist 2	Full 3	Environmentalist 4
DLP	0.00009751 (0.1674)	0.00009658 (0.1683)	0.0001313 (0.189)	0.00012826 (0.191)
DLS	1.860956 (5.88)	1.804828 (5.98)	2.112697 (7.07)	1.991418 (6.83)
DSP	-2.206553 (18.70)	-2.26989 (20.32)	-2.144618 (20.77)	-2.145823 (21.48)
DSS	-5.07985 (24.13)	-5.1418 (24.84)	-5.0254 (24.74)	-5.0233 (24.95)
DC	-4.57084 (28.13)	-4.6275 (30.02)	-4.5447 (29.93)	-4.53003 (30.42)
COST	-0.00101676 (20.47)	-0.0010044 (22.71)	-0.0010783 (25.22)	-0.0010507 (26.25)
ADIST, Large	-0.00030236 (0.8577)	-0.0003796[a] (1.099)	-0.0002266 (0.665)	-0.00031988[a] (0.94)
ADIST, Small	-0.00001501 (0.061)	0.00002958[a] (0.141)	-0.00009604 (0.409)	-0.000109195[a] (0.54)
POLDIST, Large	-0.008512 (0-0337)	-0.0003796[a] --	0.066131 (0.270)	-0.00031988[a] --
POLDIST, Small	-0.0002384 (0.035)	0.00002958[a] --	-0.0011984 (0.178)	-0.000109195[a] --
AACRE(S), Large	0.00006533 (1.099)	0.00002050[a] (0.447)	...	...
AACRE(S), Small	+0.0000815 (2.155)	0.0001144[a] (3.74)	...	...
AACRE(C), Large	...	...	0.000093676 (3.847)	0.00004443[a] (2.28)
AACRE(C), Small	...	...	0.000030082 (1.602)	0.00003878[a] (2.88)
POLACRE(S), Large	-0.00063139 (1.056)	-0.00002050[a] --	...	...

Table 10.4 (continued)

Variable (S)=State, (C)=County Large=Big Boat Small=Small Boat	State Level Freshwater Acres Full 1	 Environmentalist 2	County Level Freshwater Acres Full 3	 Environmentalist 4
POLCARE(S), Small	+0.000367 (1.086)	-0.0001144[a] --	...	...
POLACRE(C), Large	...	...	-0.00124123 (2.83)	-0.00004443[a] --
POLACRE(C), Small	...	...	-0.00009239 (0.4124)	-0.00003878 --
HDD	-0.00002277 (1.22)	-0.00002633 (1.48)	-0.00004662 (2.72)	-0.00004205 (2.54)
N of OBS	7020	7020	7020	7020
Log L	-3554.54	-3556.29	-3552.18	-35457.28

Note:

a. Parameter constrained.

Table 10.5. General National Choice Models: Negative Square Hoot Transform of Acres Freshwater Per Acre Surface Area Availability Proxy (absolute value of "t" statistic) in parenthesis

Variable (S)=State, (C)=County Large=Large Boat S=Small Boat	State Distance Transform Full 5	State Distance Transform Environmentalist 6	County Distance Transform Full 7	County Distance Transform Environmentalist 8
DLP	0.0000999 (0.16)	0.0000941 (0.14)	0.00013464 (0.22)	0.0001328 (0.20)
DLS	1.384502 (4.11)	1.390924 (4.16)	1.867755 (6.46)	1.883218 (6.54)
DSP	-1.631063 (8.62)	-1.719230 (9.4377)	-2.058567 (20.24)	-2.045076 (20.61)
DSS	-4.4893 (17.39)	-4.578003 (18.53)	-4.932221 (24.37)	-4.919467 (24.45)
DC	-3.9222 (17.54)	-4.012853 (19.17)	-4.423430 (29.25)	-4.413640 (29.49)
COST	-0.0008928 (14.90)	-0.0008963 (15.12)	-0.0010199 (25.99)	-0.0010257 (24.40)
ADIST, Large	-0.0002364 (0.68)	-0.0002577[a] (0.74)	-0.0003622 (1.05)	-0.0004207[a] (1.23)
ADIST, Small	0.00008898 (0.36)	0.0000571[a] (0.26)	-0.0001180 (0.51)	-0.0000968[a] (0.48)
POLDIST, Large	-0.034399 (0.14)	-0.0002577[a] --	-0.08210 (0.32)	-0.0004207[a] --
POLDIST, Small	-0.001084 (0.16)	0.0000571[a] --	-0.0002087 (0.03)	-0.0000968[a] --
POISSON(S), Large	-0.037084 (1.95)	-0.0429711[a] (2.33)	...	...
POISSON(S), Small	-0.06591 (3.23)	-0.0502218[a] (2.92)	...	...
POISSON(C), Large	...	...	0.0008001 (0.37)	0.0004042[a] (0.17)
POISSON(C), Small	...	...	-0.0046934 (1.67)	-0.003993[a] (1.70)
POLPOTS(S), Large	-0.27614 (1.33)	-0.0429711[a] --	...	...

Table 10.5 (continued)

Variable (S)=State, (C)=County Large=Large Boat S=Small Boat	State Distance Transform		County Distance Transform	
	Full	Environmentalist	Full	Environmentalist
	5	6	7	8
POLPOIS(S), Small	0.07709 (0.96)	-0.0502218[a] --	...	...
POLPOIS(C), Large	...	...	-0.122097 (1.14)	0.0004042[a] --
POLPOIS(C), Small	...	...	0.0108634 (0.43)	-0.003993[a] --
HDD	-0.00002931 (1.64)	-0.0000320 (1.88)	-0.00004870 (2.86)	-0.0000465 (2.80)
N of OBS	7020	7020	7020	7020
Log L	-3553.90	-3556.03	-3559.89	-3560.92

Notes:
a. Restricted

Table 10.6. Restricted National Choice Models: Acres Freshwater Per Acre Surface Area Availability Measure
(absolute value of "t" statistic) in parenthesis

Variable (S)=State (C)=County	State Level Freshwater Acres Full 9	Skeptic 10	Environmentalist 11	County Level Freshwater Acre Full 12	Skeptic 13	Environmentalist 14
DLP	0.0001218	0.0001247	0.0001229	0.0001597	0.0001598	0.0001585
	(0.19)	(0.19)	(0.19)	(0.22)	(0.23)	(0.22)
DLS	1.986776	2.011477	2.027448	2.052774	2.006116	2.015642
	(6.93)	(7.19)	(7.24)	(7.25)	(7.17)	(7.20)
DSP	-2.127185	-2.142159	-2.151174	-2.158115	-2.133644	-2.138577
	(21.21)	(23.20)	(23.04)	(22.42)	(23.12)	(23.08)
DSS	-5.004243	-5.019854	-5.029469	-5.037300	-5.011302	-5.016631
	(24.74)	(25.34)	(25.32)	(25.17)	(22.30)	(25.30)
DC	-4.509563	-4.528054	-4.539538	-4.550507	-4.519191	-4.525595
	(29.33)	(30.98)	(30.85)	(30.38)	(30.91)	(30.88)
COST	-0.0010476	-0.0010536	-0.0010574	-0.0010647	-0.0010536	-0.0010558
	(27.54)	(30.08)	(29.95)	(29.11)	(29.92)	(29.91)
ADIST	-0.0001045	-0.0000909	-0.0000827	-0.0001445	-0.0001680	-0.0001623
	(0.51)	(0.49)	(0.45)	(0.73)	(0.93)	(0.91)
POLDIST	0.0000214	**0**[a]	**-0.0000827**[a]	-0.0007729	**0**[a]	**-0.0001623**[a]
	(0.00)	--	--	(0.12)	--	--
AACRE(S)	0.0000737	0.0000820	0.0000864	...	...	...
	(2.23)	(3.23)	(3.18)			
POLACRE(S)	0.0001191	**0**[a]	**-0.0000864**[a]	...	...	...
	(0.39)	--	--			
AACRE(C)	...	...	...	0.0000504	0.0000376	0.0000405
				(3.26)	(3.43)	(3.49)
POLACRE(C)	...	...	...	-0.0002274	**0**[a]	**-0.0000405**[a]
				(1.15)	--	--
HDD	-0.0000229	-0.0000245	-0.0000261	-0.0000453	-0.0000418	-3.0000423
	(1.23)	(1.36)	(1.47)	(2.67)	(2.52)	(2.56)
N of OBS	7020	7020	7020	7020	7020	7020
Log L	-3558.01	-3558.14	-3558.24	-3557.02	-3557.74	-3557.57

Note:
a. Parameter constrained.

Table 10.7. Restricted National Choice Models: Negative Square Root Transform or Acres Freshwater Per Acre Surface Area Availability Proxy
(absolute value of "t" statistic) in parenthesis

Variable		State Level Freshwater Distance Proxy		County Level Freshwater Distance Proxy		
(S)=State (C)=County	Full 15	Skeptic 16	Environmentalist 17	Full 18	Skeptic 19	Environmentalist 20
DLP	0.0001152 (0.17)	0.0001168 (0.17)	0.0001206 (0.18)	0.0001487 (0.21)	0.0001573 (0.23)	0.0001493 (0.21)
DLS	1.344182 (4.22)	1.358160 (4.31)	1.409929 (4.53)	1.887283 (6.71)	1.886559 (6.72)	1.889187 (6.73)
DSP	-1.703918 (12.52)	-1.708847 (12.74)	-1.735816 (13.28)	-2.055872 (21.77)	-2.053815 (21.94)	-2.053510 (21.94)
DSS	-4.560985 (20.35)	-4.566266 (20.50)	-4.594929 (20.84)	-4.930071 (24.75)	-4.927894 (24.79)	-4.927760 (24.79)
DC	-3.989196 (20.85)	-3.996140 (21.13)	-4.031055 (21.76)	-4.424150 (29.88)	-4.422061 (29.98)	-4.422031 (29.98)
COST	-0.0008830 (16.47)	-0.0008865 (16.92)	-0.0008996 (17.76)	-0.0010256 (28.39)	-0.0010256 (28.48)	-0.0010260 (28.55)
ADIST	-0.0000352 (0.17)	-0.0000284 (0.15)	-0.0000310 (0.16)	-0.0002050 (1.04)	-0.0001925 (1.07)	-0.0001864 (1.04)
POLDIST	0.0003227 (0.05)	0[a] --	-0.0000310[a] --	0.0012240 (0.18)	0[a] --	-0.0001864[a] --
POISSON(S)	-0.0528105 (3.62)	-0.0514959 (3.71)	-0.0466445 (3.60)	...	...	...
POLPOIS(S)	0.0234104 (0.31)	0[a] --	-0.0466545[a] --	...	...	...
POISSON (C)	...	...	...	-0.0023841 (1.32)	-0.0023779 (1.36)	-0.0023272 (1.37)
POLPOIS(C)	...	...		-0.0015744 (0.06)	0[a]	-0.0023272[a] --
HDD	-0.0000297 (1.71)	-0.0000305 (1.80)	-0.0000330 (1.96)	-0.0000474 (2.82)	-0.0000470 (2.83)	-0.0000471 (2.84)
N of OBS	7020	7020	7020	7020	7020	7020
LOG L	-3555.88	-355.98	-3556.34	-3561.95	-3562.03	-3561.98

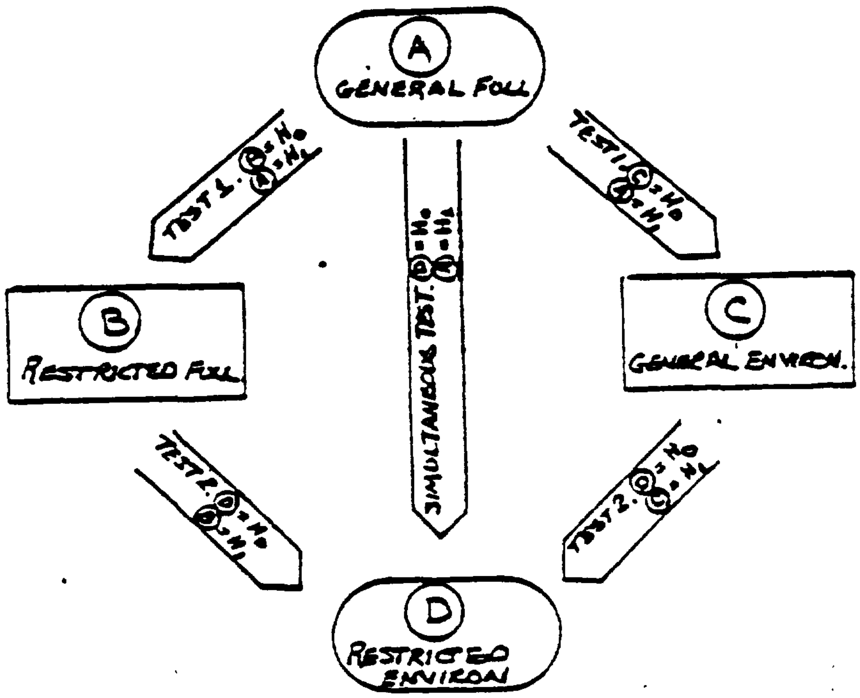
Notes:
a. Parameter constrained.

evidence elsewhere, might be something of an artifact because the fraction polluted information itself is measured at the state, not the county, level. From these results one might conjecture, that, given the data (at least if a nationally representative indirect utility function hypothesis is maintained) it is freshwater availability, not distance to Marine or Great Lakes coastline, that appears to have the stronger influence on boat ownership, and pollution hardly seems to matter.[19]

These observations raise the question of how the parameter restrictions of the models outlined in figure 10.4 can be tested statistically. This is a problem of sequential hypothesis testing and there are two routes which could be followed. Recall that we have the set of restrictions involved in the General versus Restricted specifications, and the set of restrictions involved in the Full versus Environmentalist specifications (for the moment ignoring the Skeptic model and the question of regionally differentiated versus nationally shared indirect utility functions). The testing sequence under those conditions is diamond-shaped, as indicated in figure 10.5.

One can either follow the two-step route H_1=A, H_0=B and, if the restrictions of B cannot be rejected, test H_1=B, H_0=D or go down the other side of the tree and test H_1=A, H_0=C first and, if H_0=C cannot be rejected, test H_1=C, H_0=D in a second step. Naturally, the simultaneous test of A versus D can be performed as well.

Figure 10.5. Sequential and Simultaneous Tests of Parameter Restrictions



Zero parameter restrictions implied by the null hypothesis of either B, C, or D may be tested by the likelihood ratio procedure, defining LR as $-2\ln\lambda = 2(\ln L_1 - \ln L_0)$ where λ is the ratio of the maximized likelihood function under the null hypothesis, L_0, to the unconstrained maximized likelihood function L_1. This statistic is asymptotically distributed as chi-square with n degrees of freedom, n being the number of restrictions imposed to define the null hypothesis: In performing a sequential test, to keep the significance level α of the overall family of n statements in line with the significance level of the simultaneous test of H_1=A versus H_0=D, the significance levels of the individual test statistics α_i in the sequence of tests (n=2) must be scaled such that $\alpha_i = \alpha/n$. (See Miller 1966 for a discussion of this procedure, which is based on the Bonferroni inequality). The results of the tests for the various specifications using alternative definitions of the freshwater availability measure appear in table 10.8, where the overall error rate α is set at 0.05.

The results of the sequential left-branch, right branch and simultaneous tests all lead to an inability to reject the restricted environmentalist model specification, however the freshwater availability variable is measured. Given this conclusion, how would the Restricted Skeptic model fare in position D vis-a-vis the General model in position A, assuming a common national utility function? The answer, performing a simultaneous rather than sequential test, is that the null hypothesis of the Restricted Skeptic cannot be rejected at the 5 percent level either, as shown in table 10.9. This result makes it impossible to discriminate between the non-nested Skeptic and Environmentalist specifications, a disquieting but not unexpected result. Recognizing that the benefits of water pollution control accruing to new boaters have a defensible lower bound of zero (the Skeptic models) the range of positive benefits are obtained from the various Environmentalist specifications (models 11, 14, 17 and 20) in the next section.

WELFARE ESTIMATES FROM THE ECONOMETRIC RUM'S

To derive money metric estimates of the utility change of, say, bringing all water currently unsuitable for boating for pollution-related reasons up to boatable quality, either a marginal (Eq. 30) or non-marginal (Eq. 35) compensating variation measure can be calculated. The benefit measures thus produced measure only that portion of the total benefit attributable to incremental changes in the utility boat durable ownership,

Table 10.8. Sequential Hypotheses Tests and a Simultaneous Test

I. Left Branch Sequence

1. First Test, A versus B, critical χ^2 for $\alpha = .025$, m=4 is 11.143

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres/ Acre	Model 1	Model 9	6.94	Accept H_0
County, Acres/ Acre	Model 3	Model 12	9.68	Accept H_0
State, (Acres/ Acre) $^{-1/2}$	Model 5	Model 15	3.96	Accept H_0
County, (Acres/ Acre) $^{-1/2}$	Model 7	Model 18	4.12	Accept H_0

2. Second Test, B versus D, Critical χ^2 for α = .025, m = 2 is 7.378

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres/ Acre	Model 9	Model 11	0.46	Accept H_0
County, Acres/ Acre	Model 12	Model 14	1.10	Accept H_0
State, (Acres/ Acre) $^{-1/2}$	Model 15	Model 17	0.92	Accept H_0
County, (Acres/ Acre) $^{-1/2}$	Model 18	Model 20	0.06	Accept H_0

Table 10.8 (continued)

II. Right Branch Sequence

1. First Test, A versus C, Critical χ^2 for $\alpha=.025$, m=4 is 11.143

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres/ Acre	Model 1	Model 2	3.50	Accept H_0
County, Acres/ Acre	Model 3	Model 4	10.20	Accept H_0
State, (Acres/ Acre)-1/2	Model 5	Model 6	4.26	Accept H_0
County, (Acres/ Acre)$^{-1/2}$	Model 7	Model 8	2.06	Accept H_0

2. Second Test, C versus D, Critical χ^2 for $\alpha = .025$, m = 2 is 7.378

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres Acre	Model 2	Model 11	3.90	Accept H_0
County, Acres/ Acre	Model 4	Model 14	0.58	Accept H_0
State, (Acres/ Acre)-1/2	Model 6	Model 17	2.20	Accept H_0
County, (Acres/ Acre)-1/2	Model 8	Model 20	2.46	Accept H_0

Table 10.8 (continued)

III: Simultaneous Test, A versus D, Critical χ^2 for α=0.05, m=6 is 12.592

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres Acre	Model 1	Model 11	7.40	Accept H_0
County, Acres/ Acre	Model 3	Model 14	10.78	Accept H_0
State, (acres/ Acre)$^{-1/2}$	Model 5	Model 17	4.88	Accept H_0
County, (Acres/ Acre)$^{-1/2}$	Model 7	Model 20	4.18	Accept H_0

Table 10.9. Simultaneous Test of General Full Versus Restricted Skeptic Models (Critical χ^2 for $\alpha=.05$, m=6 is 12.592)

Freshwater Availability Variable	H_1	H_0	$-2\ln\lambda$	Decision
State, Acres/ Acre	Model 1	Model 10	7.20	Accept H_0
County, Acres/ Acre	Model 3	Model 13	11.12	Accept H_0
State, (Acres/ Acre)$^{-1/2}$	Model 5	Model 16	4.16	Accept H_0
County, (Acres/ Acre)$^{-1/2}$	Model 7	Model 19	4.28	Accept H_0

not the portion attributable to increased utility of boat services among existing durable good owners.

Because such a small portion of all water is unsuitable for boating due to pollution (roughly 3 percent), the marginal and non-marginal benefit measures should be nearly equivalent.

In fact they are. To demonstrate, the pre and post policy values of the variables affected by water pollution control appear in table 10.10. Using these values, along with the sample means of the other variables, produces the non-marginalist calculations, an example of which appears in table 10.11, where model 17 is evaluated. This table demonstrates that the policy marginally increases the utility of boat ownership relative to non-ownership, which translates into a slight fall in the predicted probability of non-ownership. The benefit of making all water (fresh, marine and Great Lakes) boatable is $1.92 per household (note CV is always negative for welfare improvements, but we report the absolute value of CV, this being understood). Obviously, the non-marginalist calculations, while straightforward, are rather tedious. Performing the marginalist calculation, again for model 17, is quick and simple. The procedure is demonstrated in table 10.12, where for convenience the sample class frequencies are used in lieu of the predicted choice category probabilities. The total benefit of complete freshwater, Great Lakes and marine cleanup is again $1.92 per household, where the two components, freshwater and marine/Great Lakes, are additive due to the path independence of the compensating variation measure.

Calculations identical to those of the table 10.12 example made for the other three environmentalist models (11, 14, and 20) produce the per household and national boat ownership benefit estimates of complete cleanup

Table 10.10. Pre and Post Water Pollution Control Sample Mean Values of Variables Affected by Complete Cleanup Policy

Model[a]	Variables Affected By Policy	Sample Means of Variables: Pre Policy	Post Policy
11	AACRE(S), Large Boats	1943.37	1943.37
	AACRE(S), Small Boats	2097.78	2097.78
	POLACRE(S), Large Boats	59.21	0
	POLACRE(S), Small Boats	83.76	0
	ADIST, Large Boats	183.73	183.73
	ADIST, Small Boats	187.41	187.41
	POLDIST, Large Boats	0.12	0
	POLDIST, Small Boats	2.82	0
14	AACRE(C), Large Boats	2121.05	2121.05
	AACRE(C), Small Boats	2260.30	2260.30
	POLACRE(C), Large Boats	57.21	0
	POLACRE(C), Small Boats	83.23	0
17	POISSON(S), Large Boats	8.81	8.81
	POISSON(S), Small Boats	8.11	8.11
	POLPOIS(S), Large Boats	0.20	0
	POLPOIS(S), Small Boats	0.35	0
20	POISSON(C), Large Boats	15.81	15.81
	POISSON(C), Small Boats	14.57	14.57
	POLPOIS(C), Large Boats	0.30	0
	POLPOIS(C), Small Boats	0.61	0

Note:

a. Models 14, 17, and 20 have the same sample means for ADIST and POLDIST as those reported for Model 11, so these values are not repeated:

Table 10.11. Example Non-Marginal Welfare Change and Probability Calculation, Model 17, Environmentalist, $(\text{ACRE/ACRE})^{-1/2}$, State Data

Choice Category	Variable	Sample Mean of Variable[a]	Parameter Estimate[b]	Product of Sample Mean and Parameter[c]	Category exp(S Product) Pre and Post[c]	Predicted Probability Pre and Post[d]
NO OWN	HDD	5356.75	-0.33(E-4)	-0.176896		
				Total Pre -0.176896	0.837867	0.879765
				Post -0.176896	0.837867	0.878248
LARGE POWER	COST	3373.08	-0.90(E-3)	-0.034534		
	ADIST	183.73	-0.31(E-4)	-0.005696		
	POLDIST	0.12	-0.31(E-4)	(-0.000004)		
	POISSON(S)	8.81	-0.47(E-1)	-0.410999		
	POLPOIS(S)	0.20	-0.47(E-1)	(-0.009146)		
	DLP	...	0.12(E-3)	0.120658		
				Total Pre -3.460259	0.031422	0.032993
				Post -3.451109	0.031710	0.033238
LARGE SAIL	COST	7704.40	-0.90(E-3)	-6.931135		
	ADIST	183.73	-0.31(E-4)	-0.005696		
	POLDIST	0.12	-0.31(E-4)	(-0.000004)		
	POISSON(S)	8.81	-0.47(E-1)	-0.410999		
	POLPOIS(S)	0.20	-0.47(E-1)	(-0.009146)		
	DLS	...	1.41	1.409929		
				Total Pre -5.947050	0.002614	0.002745
				Post -5.937901	0.002638	0.002765
SMALL POWER	COST	661.29	-0.90(E-3)	-0.594922		
	ADIST	187.41	-0.31(E-4)	-0.005810		
	POLDISTD	2.82	-0.31(E-4)	(-0.000087)		
	POISSON(S)	8.11	-0.47(E-1)	-0.378162		
	POLPOIS(S)	0.20	-0.47(E-1)	(-0.016328)		
	DSP	...	-1.73	-1.735816		
				Total Pre -2.731126	0.065146	0.068404
				Post -2.714711	0.066224	0.069416
SMALL SAIL	COST	539.84	-0.90(e-3)	-0.485660		
	ADIST	187.41	-0.31(E-4)	-0.005810		
	POLDIST	2.82	-0.3l(E-4)	(-0.000087)		
	POISSON(S)	8.11	-0.47(E-1)	-0.378162		
	POLPOIS	0.20	-0.47(E-1)	(-0.016328)		
	DSS	...	-4.59	-4.594929		
				Total Pre -5.480977	0.004165	0.004373
				Post -5.464561	0.004234	0.004438
CANOE	COST	70.83	-0.90(E-3)	-0.063723		
	ADIST	187.41	-0.31(E-4)	-0.005810		
	POLDIST	2.82	-0.31(E-4)	(-0.000087)		
	POISSON(S)	8.11	-0.47(E-1)	-0.378162		
	POLPOIS(S)	0.20	-0.47(E-1)	(-0.016328)		
	DC	...	-4.03	-4.031055		
				Total Pre -4.495165	0.011163	0.011721
				Post -4.478750	0.011348	0.011895
Grand Total					Pre -0.952376	1.000000
					Post -0.954021	1.000000

Compensating Variation $(CV)^e$ = (ln 0.952376-ln 0.954021)/0.0009 = -\$1.92.

Notes:

a. From table 3.
b. From table 7. Factor in parenthesis represents 10^{-X}.
c. Components may not add to totals due to rounding. Values in parenthesis set to zero for post-policy evaluation.
d. Ratio of Category Total to Grand Total from preceding column. See formula (20) in the text.

Table 10.12. Example Marginal Welfare Change Calculation, Model 17, Environmentalist $(ACRE/ACRE)^{-1/2}$, State Data

Choice Category	Probability P_i	MRS_i[b]	Change in Proxy for Expected Distance, $\Delta\phi_i$[c]	CV = $p_i \cdot MRS_i \Delta\Phi_i$
I. FRESHWATER				
NO OWN	0.8778	0	0	0
LARGE POWER	0.0336	-51.85	0.20	-0.34
LARGE SAIL	0.0033	-51.85	0.20	-0.03
SMALL POWER	0.0691	-51.85	0.35	-1.25
SMALL SAIL	0.0044	-51.85	0.35	-0.08
CANOE	0.0118	-51.85	0.35	-0.21
SUBTOTAL				(-1.91)
II. MARINE/GREAT LAKES				
NO OWN	0.8778	0	0	0
LARGE POWER	0.0336	-0.03	0.12	-0.0001
LARGE SAIL	0.0033	-0.03	0.12	=0
SMALL POWER	0.0691	-0.03	2.82	-0.0067
SMALL SAIL	0.0044	-0.03	2.82	-0.0004
CANOE	0.0118	-0.03	2.82	-0.0012
SUBTOTAL				(-0.0084)
III. GRAND TOTAL				$-1.92

Notes:

a. Mean Sample Frequencies.

b. Marginal utility of expenditure, A, is the negative of the estimated boat cost parameter, -0.0008996 (t statistic = 17.76). Change in utility for a decrease in the travel distance proxy $(Acres/Acre)^{-1/2}$ due to pollution control is the estimated parameter on the distance proxy, -0.0466445 (t statistic = 3.60) so:

$$MRS_i = (\partial\bar{v}_i/\partial\phi_i)/\lambda = -0.0466445/0.0008896$$
$$= -51.85$$

c. Pre-policy values of pollution fraction (which differs by boat size category, 2% for large boats, 4% for small boats) times negative square root of freshwater acres per acre surface area. Sample mean values from table 10.10.

reported in table 10.13. From this table we observe that model specification differences cause a wide variation (a factor of 10 from lowest to highest) in the benefit estimates, making it difficult to choose a best estimate when all specifications are non-nested and plausible. Second, whatever the specification, it appears the benefits of marine cleanup are low relative to freshwater cleanup, being at the most 22 percent of the latter and at the least, negligible. Finally, the total magnitude of the national ownership benefit is small, which is not an unreasonable result considering the marginal nature of the improvement and the essentially footloose nature of boating. Being an inherently mobile pursuit, the ability to avoid patches of pollution is an attribute of boating by definition, perhaps diminishing the importance of pollution considerations in the purchase decision.

To sum up, unless large pollution control benefits can be uncovered which accrue to the existing universe of boat owners and boat renters over and above the benefits which accrue from the increased utility of new ownership, the overall national boating benefits of water pollution control are not likely to be as large as, say the fishing benefits. Making this conjecture even more plausible is the fact that some fishing and hunting related benefits are already captured in the totals reported above, which pertain to boat ownership for all purposes, including fishing and hunting as well as pleasure cruising.

The next chapter investigates other possible sources of benefits, particularly the benefits of cleaner water accruing to existing boat owners. The results in that chapter suggest that these sources are relatively minor, given the data employed for benefit estimation

Table 10.13. The Per Household and National New Ownership Benefits of Attaining Boatable Quality Water from 1976 Pollution Levels

Restricted Environmentalist Model	Benefits[a] Freshwater	Marine	Total
11. State, Acres/Acre			
Per Household (1976 dollars)	0.75	0.02	0.77
National (million 1976 dollars)	60.00	1.60	61.60
National (million 1983 dollars)	102.60	2.74	105.34
14. County, Acres/Acre			
Per Household (1976 dollars)	0.35	0.04	0.39
National (million 1976 dollars)	28.00	3.20	31.20
National (million 1983 dollars)	47.88	5.47	53.35
17. State, $(\text{Acres/Acre})^{-1/2}$			
Per Household (1976 dollars)	1.92	0.01	1.93
National (million 1976 dollars)	153.60	0.80	154.40
National (million 1983 dollars)	262.66	1.37	264.02
20. County $(\text{acres/Acre})^{-1/2}$			
Per Household (1976 dollars)	0.14	0.04	0.18
National (million 1976 dollars)	11.20	3.20	14.40
National (million 1983 dollars)	19.15	5.47	24.62

Note:

a. National benefits based on the product of the per household benefit and a national total of 80 million households, net of Alaska. Hawaii and the District of Columbia. (Statistical Abstract of the United States, 1984). Price index used to convert 1976 dollars to 1983 dollars is 1.71.

NOTES

1. For example, consumers may consider power boats more versatile and easier to operate than sail boats.

2. As McFadden, 1982 observes, this specification ignores the temporal dynamics of durable purchase decisions and treats historical ownership as synonymous with contemporary purchase. McFadden remarks that "this is correct only under the implausible assumption that there is a perfect rental market without transactions costs, or else that costs have not shifted or were perfectly anticipated since date of purchase", (p. 9).

3. Usually the assumption is made that the consumer cannot exhaust his income by purchasing a durable, so $p_i x_i < y$ for all i.

4. In Hanemann (1981), every consumer purchases a durable good, so the possibility represented by (4) is not covered. In this form the model is appropriate for analysis of the brand choice of a nearly universally held durable (refrigerators, for example) or a transit modal choice decision, where everyone in the relevant population must travel. So, the model is only a special case of the more general situation set out above.

5. Small and Rosen 1981 observe that as long as the Hicksian composite is perfectly divisible, the assumption that $u(\cdot)$ is strictly increasing in z and non-decreasing in x_i guarantees that the indirect utility function exists and is strictly increasing in y.

6. Note that if all of the π_i values were negative for all individuals in a large sample, all observations would exhibit durables purchase, which is the specific case analyzed by Hanemann (1982.a) and Small and Rosen (1981).

7. Intrapersonal random utility means that each individual is a rational utility maximizer given his state of mind, but the latter varies randomly from one choice situation to the next.

8. The analysis of more than two choices is a straightforward generalization. See, for example, Hensher and Johnson 1981.

9. See Henscher and Johnson 1981. Generic variables (like price in the above example) vary in level across choices but have a common parameter, while alternative-specific dummy variables correspond to the effect of a specific alternative on utility, and continuous alternative specific variables (like climate) result from the interaction of an alternative-specific dummy variable and an attribute of the individual or the environment.

10. Alaska, Hawaii and the District of Columbia were excluded from our version of the Coast Guard sample.

11. Our analysis is confined to ownership of a single durable, and thus bypasses complications of modeling multiple-unit choices. Large boat owners who also reported a small secondary purpose boat were included oh the rationale that many large boats are customarily equipped with an auxiliary dinghy or raft. One of the potential advantages of the Coast Guard survey is that it asked for the percentages of total boating time that were spent pleasure cruising and sailing, water skiing, racing, canoeing, kayaking, white water rafting, hunting and fishing. Therefore it is possible to distinguish between boat ownership whose primary purpose was the enjoyment of boating in its own right versus ownership as an input to another recreational activity such as hunting or fishing. Potentially this permits the isolation of the benefits of water quality improvements accruing solely to pleasure boating uses, avoiding double counting problems

which could occur if hunting and fishing benefits were also indirectly derived from the analysis of boat ownership for all uses. However, to do so a nested (or sequential) logit model based on a hierarchical decision tree would have to be estimated assuming the generalized extreme value (GEV) error distribution (McFadden 1982). Additionally, variables distinguishing boat ownership for pleasure versus fishing and hunting would be required to make the nested model realistic. Because of the more demanding computational and data requirements of the nested logit model we confine our analysis to the simpler conditional logit model which can be shown to be a restricted version of the former (Maddala 1983). We do not distinguish boat ownership by type of use.

12. These figures require sane adjustments before they can be used - particularly the exclusion of saline water and the addition of the acreage of small lakes. Such adjustments are discussed in Hewitt and Zimmerman 1984.

13. Residents of Alabama, Florida, Georgia, Indiana, Maryland, Massachusetts, Nevada, New York and West Virginia could not be included in our final sample due to the failure of these states to respond to that part of Dyson's survey.

14. To avoid the loss of a large portion of the sample due to the absence of marine pollution and non-pollution limitation data for Massachusetts, Maryland and New York, an average of neighboring state marine percentages was substituted. Maine's values were imputed to Massachusetts; an average of New Jersey, Delaware and Virginia's values to Maryland, and an average of Connecticut and New Jersey's values to New York. Unlike the case of freshwater missing values, this imputaition was necessary because residents

of many states have these coastlines as their ultimate destinations, and missing values here would drastically reduce our sample size.

15. For example, Illinois reports 67.5 percent of Great Lakes water unavailable for reasons other than pollution, while Indiana reports no non-pollution limitations. Yet Indiana's lake county shoreline is heavily industralized while numerous parks dot the Illinois coastline.

16. In general, suppose pollution increases the expected travel distance to a site by the relation $D_A=D_O/(1-P_M)$ where D_O is the unadjusted original distance and D_A is the pollution-adjusted distance. A first-order Taylor's series expansion of this relation around $P_M=0$ yields the approximation $D_A \approx D_O+\Delta P_M D_O$ where ΔP_M is the increment in the fraction of water polluted from a base of 0. This decomposition of the overall effect is used in specification of the full model so that the influence on choice of pollution term in the hypothesized relation $D_A=D_O/(1-P_M)$ can be tested statistically.

17. This sort of specification was employed in the fishing participation analyses of Vaughan and Russell 1982, and is perhaps more reasonable in that context than it is for boating.

18. Parameter estimates for the activity-specific dummy variables in this sort of model are not amenable to meaningful interpretation.

19. This conclusion is preliminary and can be explored further by creating activity-specific distance variables which allow each choice to react differently to distance, rather than imposing parameter equality.

APPENDIX A. EXPECTED ANNUAL COSTS OF BOAT OWNERSHIP AND OPERATION

The cost variable in our conditional logit model reflects the annual rental price of capital services, miscellaneous fees, insurance, and the average operating cost for a typical or standard year, all distinguished by category of boat owned. These components are discussed below.

CAPITAL COSTS AND SERVICE PRICES

Assume an exogenously fixed useful service life of an asset which cannot be altered by adjustments in maintenance costs per annum. Then, the rental or service price per annum of the durable consumer good can be calculated from a simplified version of the service price formula (Christensen and Jorgensen 1969). Ignoring capital appreciation/depreciation and property taxes,[1] the annual service price, S, is the sum of the cost of capital to the household, the current cost of replacement, boat registration fees, and insurance:

$$S = (r + u)A + F + I \tag{1}$$

where r is the effective after-tax rate of return, u is rate of replacement of consumer's durable, A is the capital cost of the boat, including sales tax, F is the boat registration fee and I the insurance premium.

We assume an after-tax rate of return of 8 percent for r and a replacement rate, u, of 20 percent. The latter is derived by assuming a mean useful life of 10 years (U.S. Department of Commerce 1982) and a double-declining balance depreciation schedule, which implies u = 2/10 (Christensen and Jorgensen 1969).[2]

The 1983 before-sales tax asset costs of boats of various types appear in table 10.A.1, along with their 1976 dollar equivalents, obtained by

Table 10.A.1. Before Tax Boat Costs

Boat Type	Capital Cost[a] 1983 Dollars	1976 Dollars[f]
Large (≥16') Power[b]	17,000	9,934
Large (≥16') Sail[c]	42,000	24,543
Small (16'<) Power[d]	2,800	1,636
Small (16'<) Sail	2,900	1,695
Small (16'<) Other[e]	400	234

Notes:

a. Length-weighted and ownership-weighted average of costs. Weights calculated from U.S. Department of Transportation 1978; 1983 boat costs from National-Marine Manufacturers Association 1984.

b. Includes inboard, inboard outdrive and outboard bowriders and other runabouts and inboard cabin cruisers. Cost includes motor.

c. Includes auxiliary powered sailboats.

d. Includes powered rowboats, johnboats, runabouts and other open lightweight boats.

e. Includes principally canoes, kayaks and inflatables.

f. Deflator is 1.71.

deflating the 1983 capital costs by the GNP deflator for consumer durables (U.S. Department of Commerce, Survey of Current Business, various years). The after sales tax asset costs are obtained by applying the stat&specific sales tax rates from Tax Foundation Inc., 1975 to the 1976 asset costs in table 10.A.1. Boat registration fees are state-specific and depend on the length and type of boat, and were obtained from the National Marine Manufacturer's Association. A final element of capital cost is insurance, I, which we calculated as 1.5 percent of asset cost for large power boats, 1 percent for large sail boats, 3 percent for small power boats, and 1.5 percent for small sailboats, based on interviews with a number of marine underwriters.

STANDARD OPERATING COSTS

The second element in the total annual boat price over and above the capital rental price is the expected annual running cost. Since a preponderance of the power boats in the sample (95 percent) used gasoline fuel (U.S. Department of Transportation 1978, table 30) our measure of expected running costs is based on the annual cost of fuel, assuming a standard 153 hours of operation per year, calculated for all boats from U.S. Department of Transportation 1978, tables 36, 37, and 38. The gallons of fuel consumed per boat per standard year of operation calculated from the same source are 351 for large power boats, 68 for large sailboats (most have auxiliary engines), 184 for small power boats, and 0 for small sail and other small non-power boats. These requirements multiplied by regionally differentiated fuel prices (Federal Energy Administration 1976) give the standard operating costs shown in table 10.A.2.

Table 10.A.2. Standard Annual Boat Operating Costs
(1976 dollars)

Region[a]	1976 Gasoline Cost Per Gallon	Standard Operating Cost Large Power	Large Sail	Small Power
New England	0.576	202.18	39.77	105.98
Mid Atlantic	0.599	210.25	40.73	110.22
Lower Atlantic	0.597	209.55	40.60	109.85
Mid Continent	0.589	206.04	40.05	108.38
Gulf Coast	0.562	197.26	38.22	103.41
Rocky Mountain	0.603	211.65	41.00	110.95
West Coast	0.607	213.06	41.28	111.69

Note:

a. Regional State composition as follows:

New England
Conn., Maine, Mass., N.H., R.I., Vt.

Middle Atlantic
Del., Md., N.J., N.Y., Pa.

Lower Atlantic
Fla., Ga., N.C., S.C., Va. W. Va.

Mid Continent
Ill., Ind., Lowa., Kans., Ky., Mich., Minn., Mo., Nebr., N. Dak., Ohio, Okla., S. Dak., Tenn., Wisc.

Gulf Coast
Ala., Ark., La., Miss., N. Mex, Tex.

Rocky Mountain
Colo., Idaho, Mont., Utah, Wyo.

West Coast
Ariz., Calif., Nev., Oreg., Wash.

COMPOSITE COST

An example calculation of the composite 1976 dollar boat service price and operating cost for an arbitrarily selected state, Massachusetts, is shown in table 10.A.3. Similar calculations for all state/boat categories produce the data in table 10.A.4. The composite cost is deflated by the state price index, p_z, obtained from Fuchs et. al. 1979.

Table 10.A.3. Boat Costs by Category for Massachusetts

Boat	Asset Cost[a]	Replacement Plus Interest-Cost[b]	Registration Fee	Operating Cost	Insurance Cost	Composite Cost	Undeflated Composite Cost[c]
Large Power	10,232	2865	10	202	149	3226	2976
Large Sail	25,279	7078	10	39	245	7372	6791
Small Power	1,685	472	10	106	49	637	587
Small Sail	1,746	489	0	0	25	514	474
Small Other	241	67	0	0	0	67	62

Note:

a. Includes sales tax at 3 percent.

b. From Eq. (1) above as 0.28 times Asset Cost

c. State price deflator is 1.08422.

Table 10.A.4. BOAT SERVICE PRICES FOR 1976 NORMALIZED BY HICKSIAN PRICE

STATE FIPS CODE	STATE	NORMALIZED LARGE POWER	NORMALIZED LARGE SAIL	NORMALIZED SMALL POWER	NORMALIZED SMALL SAIL	NORMALIZED SMALL NON-POWER NON-S
1	ALABAMA	3560.92	8154.76	695.13	575.40	74.66
4	ARIZONA	3575.85	8161.11	703.54	573.88	79.16
5	ARKANSAS	3648.20	8362.12	710.18	584.06	76.64
6	CALIFORNIA	3235.21	7381.24	639.82	523.56	67.60
8	COLORADO	3508.39	8003.25	691.61	563.91	73.29
9	CONNECTICUT	3878.34	6579.08	563.97	458.11	60.20
10	DELAWARE	3190.02	7247.34	630.13	506.25	66.34
12	FLORIDA	3781.06	8614.81	794.46	601.53	78.98
13	GEORGIA	3573.94	8153.40	701.27	572.44	74.64
16	IDAHO	3760.18	8579.79	738.44	599.12	78.62
17	ILLINOIS	3199.66	7318.05	625.91	512.85	67.07
18	INDIANA	3418.46	7811.58	669.70	545.99	71.62
19	IOWA	3226.33	7370.00	634.40	519.36	70.50
20	KANSAS	3191.73	7293.60	626.01	512.13	66.82
21	KENTUCKY	3313.41	7552.56	653.39	527.35	69.27
22	LOUISIANA	3527.64	8083.01	687.30	564.64	74.10
23	MAINE	3927.26	8994.60	766.37	621.11	82.50
24	MARYLAND	3187.17	7258.02	625.61	506.79	66.54
25	MASSACHUTSETT	2975.55	6790.87	587.41	474.32	62.21
26	MICHIGAN	3209.74	7326.22	626.60	512.13	67.01
27	MINNESOTA	3280.96	7497.93	642.91	525.46	68.68
28	MISSISSIPPI	3526.39	8091.33	685.10	565.10	74.23
29	MISSOURI	3360.05	7600.00	657.90	537.97	70.37
30	MONTANA	3603.82	8207.56	709.84	573.30	75.12
31	NEBRASKA	3222.64	7338.06	534.74	512.50	67.24
32	NEVADA	4014.56	9141.00	794.45	638.28	83.76
33	NEW HAMPSHIRE	3918.84	8907.05	773.69	622.32	81.55
34	NEW JERSEY	2850.37	6503.87	559.26	458.43	59.53
35	NEW MEXICO	3857.35	8830.55	757.45	627.07	80.85
34	NEW YORK	2554.37	5834.86	499.80	407.41	53.49
37	NORTH CAROLINA	3449.96	7873.98	679.78	549.45	72.10
38	NORTH DAKOTA	3784.76	8645.95	739.68	603.75	79.27
39	OHIO	3191.58	7265.40	627.00	514.89	70.90
40	OKLAHOMA	3574.13	8185.69	696.33	571.26	75.40
41	OREGON	3784.01	8601.07	750.84	608.29	71.61
42	PENNSYLVANIA	3202.17	7315.79	627.03	510.72	67.11
44	RHODE ISLAND	3022.33	6917.83	588.50	483.08	63.45
45	SOUTH CAROLINA	3364.97	7684.86	660.31	536.60	70.45
46	SOUTH DAKOTA	3394.39	7758.17	665.12	541.76	71.13
47	TENNESSEE	3768.52	8607.17	738.65	605.07	78.81
48	TEXAS	3704.65	6475.13	724.14	591.97	77.72
49	UTAH	3822.54	8726.06	752.48	614.72	79.94
50	VERMONT	3802.11	8690.75	742.90	607.01	79.66
51	VIRGINIA	3180.72	7258.19	625.28	506.86	66.51
53	WASHINGTON	3345.27	7644.33	643.47	524.65	68.90
54	WEST VIRGINIA	3335.70	7608.69	656.86	531.33	69.73
55	WISCONSIN	3071.01	7021.08	601.51	493.07	64.33
56	WYOMING	3245.79	7399.19	639.84	516.68	67.80

NOTES

1. In many states, boat registration fees are levied in lieu of personal property tax (National Marine Manufacturers Association 1984).

2. Although a priori large boats might be expected to have longer lives than small boats, the median age of boats owned does not vary appreciably by size (U.S. Department of Transportation 1978).

REFERENCES

Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey," Journal of Economic Literature, vol. 19 (December) pp. 1483-1536.

Ben-Akiva, Moshe and Steven R. Lerman. 1979: "Disaggregate Travel and Mobility-Choice Models and Measures of Accessibility," in David A. Hensher and Peter R. Stopher, eds., Behavioral Travel Modeling, (London: Cromm-Helm) pp. 654-679.

Christensen, Laurits R. and Dale W. Jorgensen. 1969. "The Measurement of U.S. Real Capital Input, 1929-1967," Review of Income and Wealth, vol. 15, no. 4, pp. 293-320.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica vol. 39 (September) pp. 829-844.

Deaton, Angus and John Muellbauer. 1980. Economics and Consumer Behavior (Cambridge: Cambridge University Press) Ch. 10.

Federal Energy Administration, National Energy Information Center. 1976. Monthly Energy Review (August), p. 55.

Fuchs, Victor R., Robert T. Michael, and Sharon R. Scott. 1979. "A State Price Index," National Bureau of Economic Research Paper No. 320.

Hanemann, W. Michael. 1981. "Applied Welfare Analysis with Quantal Choice Models," Working Paper No. 173, University of California-Berkeley Department of Agricultural and Resource Economics (June).

_____. 1982a. "Quality and Demand Analysis," in Gordon C. Rausser, ed. New Directions in Econometric Modeling and Forecasting in U.S. Agriculture (New York: Elsevier North Holland) pp. 55-98.

_____, 1982b. "Applied Welfare Analysis with Qualitative Response Models," Working Paper No. 241, University of California-Berkeley Department of Agricultural and Resource Economics.

_____. 1983. "Marginal Welfare Measures for Discrete Choice Models," Economics Letters, vol. 13, pp. 129-136.

_____. 1984. "Discrete/Continuous Models of Consumer Demand," Econometrica, vol. 52, no. 3 (May), pp. 541-561.

Hartman, Raymond. 1982. "The Appropriateness of Conditional Logit for the Modeling of Residential Fuel Choice," Land Economics, vol. 58, no. 4 (November), pp. 478-487.

Hensher, David A. and Lester W. Johnson. 1981. Applied Discrete-Choice Modeling. (New York: John Wiley and Sons) Ch. 2, 3.

_____. 1984. “Achieving Representativeness of the Observable Component of the Indirect Utility Function in Logit Choice Models: An Empirical Revelation,” Journal of Business, vol. 57 (April), pp. 265-280.

Hewitt, Julie A. and Kerry Zimmerman. 1984. “Recreation Supply Variables Data Base,” Unpublished Working Paper, Resources for the Future.

Johnson, Normal L. and Samuel Kotz. 1970. Continuous Univariate Distributions (New York: John Wiley and Sons) Chapter 21.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics, (New York: Cambridge University Press).

Maler, Karl-Goran, 1974. Environmental Economics (Baltimore: John Hopkins University Press for Resources for the Future).

McFadden, Daniel. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior,” in Frontiers in Economics, P. Zarembka, ed., (New York: Academic Press) pp. 105-142.

_____. and Fred Reid. 1975. “Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models," Transportation Research Record, no. 534, pp. 24-37.

_____. 1981. “Econometric Models of Probabilistic Choice," in Structural Analysis of Discrete Data with Econometric Applications, edited by Charles F. Manski and Daniel McFadden (Cambridge: MIT Press).

_____. 1982. “Qualitative Response Models," in Werner Hildenbrand, ed., Advances in Econometrics (London: Cambridge University Press) pp. 1-37.

McKenzie, George W. 1983. Measuring Economic Welfare: New Methods (New York: Cambridge University Press) 187 pp.

Miller, R. G. 1966. Simultaneous Statistical Inference (New York: McGraw Hill).,

Mitchell, Robert Cameron and Richard T. Carson. 1981. An Experiment in Determining Willingness to Pay for National Water Quality Improvements. A Report to U. S. Environmental Protection Agency (Washington, D.C. : Resources for the Future).

National Marine Manufacturers Association. 1934: Boating 1983 (Chicago: National Marine Manufacturers Association

Small, Kenneth A. and Harvey S. Rosen. 1981. “Applied Welfare Economics with Discrete Choice Models,” Econometrica, vol. 49, no. 1 (January), pp. 105-130.

Tax Foundation, Inc. 1975. Facts and Figures on Government Finance (New York: Tax Foundation) Table 160.

U.S. Department of Commerce 1979. "User Documentation for the DMA Area Resource File,” (Springfield, Va: NTIS) Document HRP-0901718.

U.S. Department of Commerce Bureau of the Census, Geography Division. 1981. 1980 State/County Area Measurement (unpublished data).

U.S. Department of Commerce, Bureau of Economic Analysis. 1982. Fixed Reproducible Tangible Wealth in the United State, 1925-1979 (Washington, D.C.: GPO) Table B.

U.S. Department of Transportation, United States Coast Guard. 1978. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey, Washington, D.C.

Vaughan, William J. and Clifford S. Russell. 1982. Freshwater Recreational Fishing, The National Benefits of Water Pollution Control. (Washington, D.C.: Resources for the Future)

_____, _____ and Julie A. Hewitt. 1984. "Pitfalls in Applied Welfare Analysis with Recreation Participation Models, RFF Quality of the Environment Discussion Paper QE85-03.

Chapter 11

BOATING: A RECURSIVE MODEL OF PARTICIPATION INTENSITY AND ITS SENSITIVITY TO POLLUTION

It has become common in studies of the demand for marketed or non-marketed goods (eg. Duan, et. al. 1983, 1984; McDonald and Moffitt, 1980; Thraen, et. al. 1978, Ziemer, et. al. 1982) to recognize that changes in prices or other exogenous variables have a twofold effect on quantity demanded. The first involves changes in the number of individuals participating in the market and the second involves changes in quantity consumed among those individuals actively participating in the market prior to the exogenous change.

In the recreation context, focusing on the service flow output of the experience (recreation days) rather than the capital and operating inputs, the basic notion is that behavior is modeled in two steps (even though the consumer's decisions may be simultaneous). The first step is the decision to participate in the activity (i.e., have positive recreation days) while the second is the decision on what level of service flow outputs (days) to enjoy, conditional on a decision being made to participate.

When using Cragg hurdles model of recreation participation the first decision is informally referred to as the participation probability step, and the second the participation intensity step. Fundamental to such models is the idea (Duan, et. al. 1984) that the expected number of recreation days for any individual, i, in the population $E(q_i)$ is the product of the probability that the individual will participate, $P(q_i>0)$ multiplied by the expected level of activity conditional on a decision to participate having been made, $E(q_i|q_i>0)$. So, in briefer notation, for any representative individual i, the unconditional expected quantity consumed $q_i^* = E(q_i)$ is:

$$q_i^* = PR_i \times \bar{q}_i = f(\bar{q}_i PR_i) \qquad (1)$$

where $PR_i = P(q_i>0)$ and $\bar{q}_i = E(q_i | q_i>0)$ are functions h(0) and g(of a set of exogenous variables, including most importantly trip cost or a proxy thereto, which is affected (reduced) by a pollution control policy. Assuming the function $f(\bar{q}_i, PR_i) = g(x)h(x)$ is continuously differentiable with respect to the cost (or availability proxy) variable x, the first differential (ie., the first term of the Taylor's series expansion of the function) can be employed as an approximation (Allen 1967) to the change in q_i^* induced by a change in x from x^0 to $x^1<x^0$:

$$\begin{aligned} \Delta q_i^* &= f(\bar{q}_i^1, PR_i^1) - f(\bar{q}_i^0, PR_i^0) \qquad (2) \\ &= [g(x^1)h(x^1)] - [g(x^0)h(x^0)] \\ &\approx g'(x)h(x)(x^1-x^0) + h'(x)g(x)(x^1-x^0) \end{aligned}$$

where primes denote partial derivatives with respect to x and by assumption $g'(x)$ and $h'(x)$ are negative, evaluated at the point of expansion, x^0. Re-writing (2) by substituting $\Delta\bar{q}_i$ for the term $g'(x)(x^1-x^0)$ and ΔPR_i for the term $h'(x)(x^1-x^1)$ the change in the expected number of recreation trips due to the policy is approximately:

$$\Delta q_i^* \approx \Delta\bar{q}_i\, PR_i + \Delta PR_i\, \bar{q}_i \qquad (3)$$

Equation 3 is an absolute version of the elasticity formula (5) in Thraen et. al. 1978. It shows that the expected quantity adjustment to a change in trip cost for any individual is composed (approximately) of two components: the change in the expected quantity consumed conditional on consumption multiplied by the initial probability of consumption plus the change in the probability of consumption multiplied by the initial quantity that would be consumed given positive consumption. The aggregate analogue to this individual result taken over entire population of N persons of whom n initially are consumers, using $n/N \approx PR_i$ in (3) and multiplying through by N is:

$$\Delta Q^* \cong N\left(\frac{\sum \Delta q_i}{N}\frac{n}{N} + \frac{\Delta n}{N}\frac{\sum q_i}{N}\right) = \Delta\bar{q}n + \Delta n\bar{q} \qquad (4)$$

where $\frac{\sum \Delta q_i}{N}$ and $\frac{\sum q_i}{N}$ are average rates, $\Delta\bar{q}$ and $\bar{q}$ respectively.

So, the total quantity adjustment is approximately equal to the increment in the quantity demanded by individuals who initially were in the market (the subject of this chapter) plus the amount demanded by new entrants (the subject of the preceding chapter).[1]

Equation (4) lies at the heart of the recreation participation equation approach to obtaining an estimate of the change in the number of recreation occasions (days or, more properly, trips) under a policy of recreation resource augmentation (or more specifically water pollution control). Under the assumption of a constant average consumer's surplus per occasion, the quantity change in (4) can be monetized to obtain a monetary measure of benefit.[2]

Accepting the constant average value assumption, the aggregate monetary analogue to (4) in service flow space, based on the participation method, is, representing value per household per trip as v:

$$B_T = B_1 + B_2 \qquad (5)$$

where $B_T = v(\Delta Q^*)$, $B_1 = v(\Delta\bar{q}n)$ and $B_2 = v(\Delta n\bar{q})$.

The estimation of B_1 is the subject of this chapter. An estimate of B_2 was obtained in the preceding chapter via the indirect utility function for durables ownership (assuming a standard annual operating rate $\bar{q}$). In the next section we perform a plausibility check on the results of the previous chapter using the approximation for B_2 given in Eq. (5) above, before proceeding to the estimation of B_1.

A PLAUSIBILITY CHECK ON THE RUM MODEL'S WELFARE ESTIMATES

To estimate B_2 under the participation approach of Equation (5), we need to know three pieces of information: the change in the number of boating households due to the policy (Δn), the number of recreation occasions (trips) per household per year prior to the policy ($\bar{q}$), and the average consumer's surplus per household per trip.

Ignoring the boat rental market for the moment, assume the probability of boat ownership and the probability of positive boat recreation days are identical (ie: no boat owners buy but fail to use their boats at least once in a season). Then the conditional logit models of the prior chapter can be used to produce estimates of Δn, since Δn is equal to the number of households N multiplied by the change in the probability of ownership (here, participation) predicted by the logit models, evaluated at the means. The results are shown in table 11-1.

Table 11.1. Estimated Changes in Boating Participation

	Conditional Logit Model			
	11	14	17	20
ΔPR^a	0.00072	0.00032	0.00152	0.00017
N^b	80	80	80	80
$\Delta n=\Delta PR(N)^c$	0.0576	0.0256	0.1216	0.0136

Notes:

a. Change in fraction of boat ownership (ie: participation). Predicted from evaluation of Logit models under pre and post policy conditions.

b. Number of households in U.S., in millions.

c. Change in number of boat owning (participating) households, in millions.

The other two pieces of information, $\bar{q}$ and v, can be derived independently of the information used to produce the logit models in the previous chapter. First, all boat owning households in the Coast Guard survey were asked for the average number of times per month their boat was used, and the number of months of use over the year, the product being a proxy for trips per year. Moreover, all boat owning households who trailered their boat (63 percent of all owners) were asked for the average round trip miles travelled per outing. With this information on the trips per year taken by trailering households and their round-trip travel distances a conditional semilog trips demand equation can be estimated to obtain a relation estimating $E(q_i|q_i>0)$ as in (1) above; and we do so below. Additionally, the same relation can be used to produce an estimate of v. Anticipating the results discussed in detail below where the trips demand equation is estimated, several point estimates of v, differing with estimation procedure (OLS, Robust Regression) employed, are reported in table 11-2.

Table 11.2. Values Per Boating Day

	Average Surplus Per Trip[b]	
Estimation Method	$ Per Household	$Per Person
I. OLS, Full Sample	121	38
II. Robust Regression, Full Sample[a]	103	32

Notes:

a. Method II is Tukey’s biweight procedure which dampens the influence of outliers. See estimation results section of this chapter.

b. Assumed travel cost of 10 cents per mile.

Obtaining $\bar{q}$ is more straightforward. The mean number of boating trips per household for all boat-trailering households in the Coast Guard survey is 25 and a more robust measure of central tendency, the median, is 15.[3]

With information on the number of new households attracted to boat ownership because of a pollution control policy (in this case complete cleanup), their typical days of boating per year, and the average consumer's surplus per day, a plausibility check on the RUM estimates of the preceding chapter can be made. For example, Logit Model 11 predicts 0.0576 million new households due to pollution control, so the product of this value, the median days (15) of boating and a robust measure of consumer's surplus yields an estimate of total benefit component B_2:

$$0.0576 \text{ million households} \times 15 \frac{\text{days}}{\text{household}} \times 103 \frac{\$}{\text{day}} = \$88.99 \text{ million.}$$

Similar calculations for a the other models of the preceding chapter using either the median (15) or mean (25) days of annual participation produces the comparison of the RUM and participation estimates of the B_2 component of the national benefits of bringing all marine and freshwater up to boatable quality shown in table 11-3.

Table 11.3. Estimates of B_2 from RUM Models Versus Participation Approximations (million dollars)

	Conditional Logit Model			
	11	14	17	20
RUM Compensating Variation (1976$)	61.60	31.20	154.40	14.40
Participation Approx. using Median Days[a]	88.99	39.55	187.87	21.01
RUM Compensating Variation, (1983$)	105.34	53.35	264.02	24.62
Participation Approx. using Mean Days[a]	148.32	65.92	312.35	35.02

Note:
a. Uses average surplus values from robust regression reported in table 11-2, which are approximately in 1980 dollars.

While the RUM compensating variation estimates are exact theoretical measures of welfare change, the participation method estimates are approximate in more ways than just being Marshallian surpluses and thus expected to overstate the benefit of pollution control (whose effect is analogous to a price decrease). First, estimates of travel cost per mile, especially for trailered vehicles, are difficult to pin down in terms of any particular year. Indeed, travel cost itself is unlikely to be constant across individuals in cross section, and need not move proportional to the general price index. The value of 10¢ per mile used to construct table 11-2 is arbitrary, and can be best regarded as being mid-way between 1976 and 1983 costs.[4] (While 106 per mile may seem low for trailering, higher costs per mile produce implausibly high average values). Second, while the average value of $103 per household per day is our preferred measure based on robust regression procedures, the OLS values from table 11-2 cannot be ruled out. Finally, using the strict expected value (the mean) of days per year rather than a more robust measure of central tendency downweighting the importance of extreme observations (which may in fact represent erroneous survey information) has a large impact on the resultant estimate of B_2.

A mixed set of conclusions can be drawn from this plausibility check. Which are emphasized depends on one's subjective point of view. Those experienced in benefit estimation in the non-marketed goods context, which is a notoriously uncertain enterprise, may find the comparisons in table 11-3 mildly heartening. At least it is possible to find a set of assumptions which together yield close concert between the RUM compensating variation measure of welfare change and the participation approximation measure. Yet from another point of view that is just the problem--the set of assumptions (especially for average value), proper for the participation approximation method can never be determined with certainty.

In the next section we sketch the structure of a recursive model of boating intensity before moving on to discuss the data and empirical analysis giving rise to the values reported in table 11-2. Finally, alternative estimates of welfare component B_1 are presented, along with a final range of values for B_T representing the sum of B_1 estimated below and B_2 discussed above and in the preceding chapter.

A RECURSIVE TWO-EQUATION MODEL OF BOATING TRIP DEMAND AND TRAVEL DISTANCE

As discussed in chapters 3 and 4 above, in a 1971 paper, Cragg proposed a set of models for situations in which an economic agent makes two (simultaneous) decisions. A dichotomous decision is made about whether or not to engage in some activity. Conditional on an affirmative for this decision, a decision is made regarding how much of the activity to pursue.

While Cragg proposed several models, the one employed here involves a situation where the quantity of boating trips per year in the second-stage decision is defined only for positive real numbers, given the first stage decision to participate as modeled in the preceding chapter. For convenience it is assumed that the conditional density of the logarithms of the positive realizations is normal.[5]

The general specification of the boating trips demand equation estimated below is of the standard sort found in the literature, involving the inclusion of travel distance, individual socioeconomic characteristics and climatic influences. (See, for example, Ziemer and Musser 1979). There is, however, the question of how to incorporate environmental pollution.

One answer would be to add resource availability variables as additional regressors in the trips demand equation, as proxies, however vague, for environmental “characteristics”. Under this specification a pollution control policy, by augmenting the resource availability, would lead to a (presumably rightward) shift in the conditional trips demand function. This this procedure is defensible when dealing with the characteristics of specific sites and the trips thereto (eg: Wennergren, et. al. 1975, Vaughan and Russell 1982).

Yet unless such resource availabilities directly affect the consumer’s utility function, (as in Bouwes and Schneider 1979) it is hard to rationalize this specification when the demand equation refers not to visits to particular sites with specific water quality characteristics but instead to the demand for the activity, whatever the sites visited or their characteristics--that is, a demand curve for the “whole experience” (Sinden and Worrell 1979). In this broader context it is simpler to hypothesize that the observed distance travelled on an average or typical trip is itself a function of the availability of fresh and marine recreational water availability, following the argument in chapter 2 above for participation analysis generally. Here, increased availability due to pollution control decreases expected travel distance, and hence marginal trip cost, without influencing the demand relation. This alternative view and its measure of Marshallian surplus is contrasted to the demand-shifting model in Panels A and B of figure 11. 1.

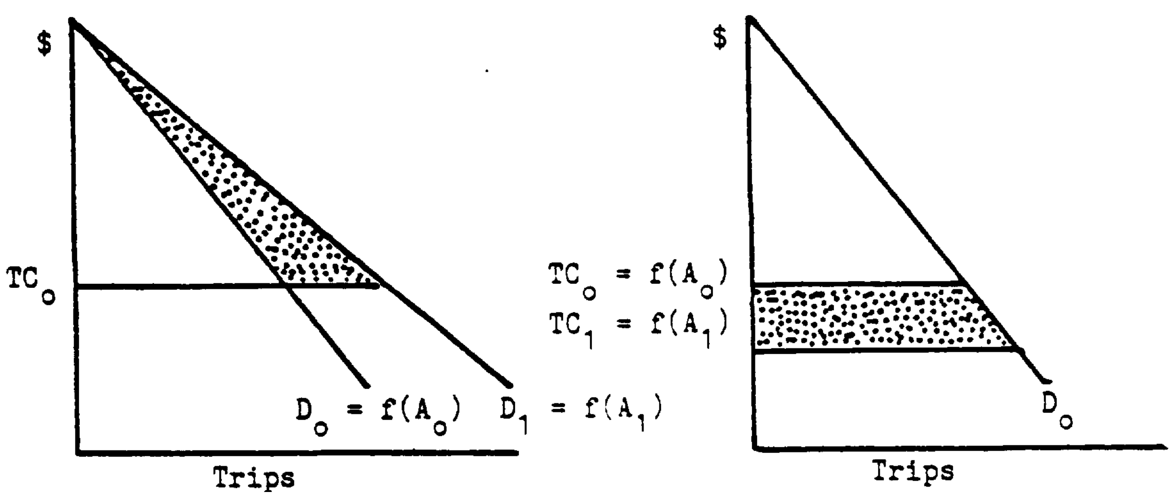
The model depicted graphically in Panel B of the figure requires, in addition to a statistically estimated trips demand function, a distance per trip (trip cost) function. So, the general recursive model involves first the decision on how far to travel per trip (D) and then, how many trips to take (Q), given a decision has been made to boat:

$$D = f(A,M) \tag{6}$$

$$Q = f(D,S) \tag{7}$$

where A represents freshwater availability, net of pollution and other limitations, M represents distance to marine or Great Lakes water, again corrected for pollution and other limitations, and S represents socioeconomic and other demand determinants.

Figure 11.1. Alternative Models Reflecting Resource Availability in Consumer Decisions



Panel A. Demand as a Function of Availability, A.

Panel B. Marginal Trip Cost as a Function of Availability, A.

DATA AND ESTIMATION ISSUES

While the general recursive structure outlined above may be plausible, several practical estimation issues must be addressed when the Coast Guard data set, with its inherent limitations, is confronted with particular specifications of (6) and (7).

The Travel Distance Function

Individuals who trailered their boats in the Coast Guard survey were asked "About how many miles, round trip, do you normally trailer or carry your boat on each outing?" While the response to this inquiry can be used as a dependent variable in the travel distance function, the form of that function needs to be specified.

Obviously the survey's trailer miles question requires the respondent to estimate sane sort of average, or typical, distance figure. Thus the response can be regarded as a probability weighted average of the respondent's expected travel distance to (presumably) the nearest marine/Great Lakes site (D_M) and the nearest freshwater site (D_F), recognizing that he may choose to recreate in both fresh and saltwater in the same aeason.[6] Letting Π_F and Π_M represent the respective probabilities that a trip was taken to marine/Great Lakes or freshwater, the general relationship determining the average miles travelled response, $\overline{D}$, is:

$$\overline{D} = \Pi_F D_F + \Pi_M D_M \qquad (8)$$

This very general notion can be represented as an estimable function in a variety of ways. The vital elements in any particular choice are first to select specific representations for Π_F and $\Pi_M = 1 - \Pi_F$ which are defined over the 0-1 interval, and second to define an estimable function whose dependent variable is logrithmic, in keeping with the Cragg's hurdles specification (Cragg 1971). This second element assumes normality or the error term and is a matter of convenience more than anything else, being a justification for not becoming involved in complex truncated regression estimation of an intrinsically nonlinear model.

A reasonable hypothesis is that Π_F and Π_M are functions of the <u>ratio</u> of the expected freshwater and marine/Great Lakes distances. With the above consideration in mind a convenient specification of the Π_F and Π_M functions which is parsimonious in parameters and confined to the 0-1 interval (Daniel and Wood 1980) is:

$$\Pi_F = e^{b_0 (D_F/D_M)} \tag{9a}$$

$$\Pi_M = 1-e^{b_0 (D_F/D_M)} \tag{9b}$$

where $b_0 < 0$.

Substituting (9.a) and (9.b) in (8) under the assumption that the error term ε_i is lognormally distributed ($\varepsilon_i \sim N(0, \sigma^2)$) yields:

$$\overline{D} - \{[\exp(b_0(D_F/D_M))]D_F + [1-\exp(b_0(D_F/D_M))]D_M\} \exp \varepsilon_i \tag{10}$$

Expressing (10) in logrithmic form yields an inherently nonlinear model whose error term is additive and normally distributed:

$$\ln\overline{D}=\ln\{[\exp(b_0(D_F/D_M))]D_F + [1-\exp(b_0(D_F/D_M))]D_M\} + \varepsilon_i \tag{11}$$

The model in (11) can be estimated using a nonlinear least squares algorithm,[7] if D_F is known a priori.

More generally, the expected two-way distance to the closest freshwater site, D_F, is proportional (Vaughan and Russell 1984) to the negative square root of the square miles of freshwater per square mile of surface area, (call it A), but itself is not directly observed. While acreage is a directly measurable quantity the factor of proportionally is not, and must be estimated since equation 11 is nonlinear.[8] Letting b_1 represent the factor of proportionality (11) can be rewritten as the trailer miles equation specification to be estimated[9], where a priori $b_0<0$, $b_1>0$:

$$\ln\bar{D} = \ln\{[\exp(b_0 b_1(A/D_M))]b_1 A + [1-\exp(b_0 b_1(A/D_M))]D_M\} + \varepsilon_i \qquad (12)$$

The variable descriptions and sample means for the variables in trailer miles equation 12, and the trip equation discussed subsequently appear in table 11.4.

The Trips Demand Equation

Utility theory suggests that any trips demand equation should ideally be specified to include as explanatory variables the travel costs of the trip, the scarcity value of time, and income (McConnell 1975). Yet many data sets preclude the possibility of valuing time cost as some fraction of the individual's wage rate because of insufficient information. In lieu of this preferred procedure, investigators often include travel cost and distance as separate regressors in the same model, the latter serving as a proxy for the opportunity cost of time. Yet since travel cost is usually constructed from distance, collinearity is often the result (Wetzstein and McNeely 1980).

Our Coast Guard data set contains no information on income or wage rates, so it is impossible to construct even a proxy for time coat. And, with our data, severe collinearity would be introduced by including both the product of distance and an (assumed) cost per mile and distance alone in the same estimating equation. Thus, we omit the scarcity value of time from consideration, recognizing the potential biases introduced by this data deficiency (ie., the true average surplus may be underestimated).

Table 11.4. Variables Used in Trailer Miles and Boating Trips Equations

Variable Name	Description[a]	Model[b]	Sample Mean[c]
I. Dependent			
MILES	Round-trip trailer miles travelled on a typical occasion	n.a.	67.99
TRIPS	Boating occasions per household per annum. Constructed as the product of occasions per month and months of boating per year.	n.a.	25.35
LNMILES	Logarithm of MILES	I	3.31
LNTRIPS	Logarithm of TRIPS	II	2.66
II. Independent			
ADIST	Distance to nearest available marine or Great Lakes coast adjusted for non-pollution limitations but assuming zero pollution.	Post-Policy Evaluation	195.56
DM	Distance to nearest available marine or Great Lakes coast, adjusted for pollution and non-pollution limitations (ie: ADIST divided by 1 minus the fraction polluted).	I	197.94
POISSON(S)	Freshwater acreage, adjusted for reasons other than pollution, divided by total state (S) surface acreage, and raised to the -1/2 power.	Post-Policy Evaluation	8.06
POISSON(C)	Constructed like POISSON (S), using county rather than state area measures.	Post-Policy Evaluation	13.20
A(S)	State freshwater acreage adjusted for pollution and non-pollution limitations, divided by total surface acreage and raised to -1/2 power.	I	8.22

II. Independent (continued)

Variable Name	Description[a]	Model[b]	Sample[c] Mean
A(C)	Constructed like A(S), using county rather than state area measures.	I	13.46
RATIO(S)	Ratio of A(S) to DM	I	0.40
RATIO(C)	Ratio of A(C) to DM	I	0.66
MILES	Round trip trailer miles travelled on a typical occasion.	II.1	67.99
PERSONS	Number of persons carried aboard the boat in a typical outing.	II	3.18
EMPLOY	Equal to 1 if primary boat owner-operator is employed, 0 otherwise.	II	0.809
HIGH	Equal to 1 if primary boat operator is high school graduate, zero other wise.	II	0.595
COLL	Equal to 1 if primary boat operator has college degree, zero otherwise.	II	0.172
PLEASURE	Percent of total boating time spent over the year in pleasure (ie: non-hunting and non-fishing) uses.	II	34.67
HDD	Annual heating degree days, a proxy for season length.	II	5140.84
LENGTH	Length of primary boat.	II	15.41
GASID	Regional gasoline price index.	n.a.	n.a.
PHICKS	State price index.	n.a.	n.a.
COST	Normalized travel cost per trip. Equal to MILES (GASID/PHICKS) multiplied by an assumed $0.10 per mile.	II.2	7.13

Notes:

a. All variables are taken directly from the Coast Guard survey except ADIST, DM, POISSON(S), POISSON(C), A(S), A(C), HDD whose sources are the same as those reported in the preceding chapter.

b. Model I is the trailer miles equation defined in Eq. 12 of the text, with a sample size of 1618. Model II is the trips equation and has two variants: II.1 using travel distance (MILES) and II.2 using travel cost. Both II.1 and II.2 share the same independent variables apart from travel cost, and are estimated from a sample of 1404 observations. Intermediate variables employed in the construction of final variables used in estimation are designated as n.a. in the “model” column.

Two alternatives are left. The first is to include distance directly as a regressor. The average surplus per household per occasion with this model is just the reciprocal of the parameter estimate on distance after division by an assumed cost per mile.[10] It is therefore unnecessary to monetize distance in this framework.

A second alternative would be to recognize that in a national cross section it is unreasonable to assume a constant coat per mile and to ignore the relative price of all other commodities, even if the operating characteristics of the vehicles used are assumed to be the same, and the scarcity value of time is ignored. In this second specification, which is consistent with demand theory, the relevant price regressor is travel cost, deflated by an index of the Hicksian composite commodity. Travel cost itself varies due to regional variation in gasoline prices.

However the travel cost variable is specified, the model should include income and climatic variables which influence the number of boating trips taken per household per year. The variables selected are shown in table 11.4. As the Coast Guard survey contains no income information, we represent it with dummy variables reflecting the employment status and educational attainment (EMPLOY, HIGH, COLL in table 11.4) of the primary boat operator. Additionally, the number of cold weather days may limit the intensity of boating activity, and we represent these with heating degree days (HDD).

Finally, the number of persons per trip and the length of the boat used could conceivably influence boating intensity (PERSONS, LENGTH). The two competing specifications of the trips demand function relate the logarithm of the number of trips per household per year to either distance (MILES) or relative trip cost (COST), along with the PERSONS, EMPLOY, HIGH, COLL, PLEASURE, HDD and LENGTH variables of table 11.4. Again consistent with the Cragg hurdles specification, the assumption of a normally distributed error term for the semi-logrithmic trips intensity equation is invoked.

The next section presents the parameter estimates for the trailer miles and trips equations discussed above. Before turning to those results, it is important to note that the Coast Guard data set contains data points which potentially represent severe outliers. For example, one industrious recreator reported taking 500 boating trips in 1976, each of which involved a round-trip trailering distance of 220 miles. Even with a large sample, the presence of a handful of such implausible responses could have considerable influence on the parameter estimates.[11]

There are many ways of dealing with potential outliers. One is to define a subset of the data which is "reliable" and to sequentially add observations from the unreliable subset, deciding each time whether to retain or drop each suspect observation based on a statistic computed from recursive residuals (Schweder 1976). However, this procedure is computer intensive and requires an initial decision defining a good subset of the data which presumably contains no outliers.

A legitimate alternative, recommended when bad data points are suspected, is robust regression (Hogg 1979), which is a formal procedure for attributing less weight to unusual data points than to typical points in estimation. Thus all models in the next section are estimated conventionally, giving equal weight to all observations, and also by a robust technique (Tukey’s Biweight) which tends to punish outliers harshly.[12]

Estimation, Results

The results from nonlinear least squares estimation of the intrinsically nonlinear trailer miles equations using state or county freshwater availability data appear in table 11.5. The overall fit of these models estimated from 1618 observations is quite good, confirming the link between distance actually travelled per recreation occasion and the water resource endowment (ie: availability) facing the individual, as argued theoretically in chapter 2 above.

The parameter estimates for b_0 and b_1 both have the hypothesized signs in all models.[13] Also, the relative stability of the parameter estimates across models (state versus county data) and estimation techniques (equal weight versus robust) is reassuring. The absence of large parameter changes across estimation techniques indicates the presence of very few outliers in the data.[14] This result suggests that, by and large, the responses given for distance travelled on a typical occasion in the survey are fairly reliable.

Yet the functions estimated in table 11.5 do exhibit one anomalous property. When the ratio of expected freshwater distance to marine/Great Lakes distance equals one, the probability of visiting freshwater, Π_F, is above 0.9. One might more reasonably expect Π_F to be in the neighborhood of 0.5 under these circumstances. But, the average boat size in the sample of trailered boats is below 16 feet, so it is not surprising to find a very high probability of freshwater use, given the risks of small boat operation in the marine environment.

Table 11.5. Alternative Nonlinear Trailer Miles Models (NLIN) (asymptotic standard errors in parenthesis)

	State Models		County Models	
Parameter[a]	NLIN I.1.a	Robust NLIN I.1.b[b]	NLIN I.2.a.	Robust NLIN 1.2. b[b]
b_0	-0.0622	-0.0664	-0.0667	-0.0651
	(0.0163)	(0.0141)	(0.0091)	(0.0078)
b_1	3.6471	3.8845	3.0147	3.1879
	(0.1442)	(0.1411)	(0.1230)	(011179)
Mean Square Error	2.160	1.633	2.280	1.703
Root Mean Square Error	1.470	1.278	1.510	1.301
R^2	0.84	0.86	0.83	0.86

Notes:

a. The parameter b_1 represents the factor of proportionality converting the negative square root of the ratio of freshwater to total surface area to a miles measure.

b. Tukey's biweight method estimated by iteratively reweighted nonlinear least squares. Robust estimates of σ required to construct weighting function formed from ordinary NLIN residuals, u_i, as $\hat{\sigma}=1.48$ (median$|u_i$ - median $u_i|$), following Holland and Welsch 1977. Respective initial robust estimates of σ are 1.430 for the state model and 1.462 for the county model.

Since Π_M is so small, the policy implication of the models in table 11.5 is that most of the effect of pollution control will be transmitted to a change in trailer miles (and hence travel cost) via freshwater, not marine, water quality improvement. Also, since marine pollution is negligible in the pre-policy situation this phenomena may tend to overstate the cost reduction from pollution control, and hence overstate the benefits therefrom. Lacking travel distance information for trips by the owners of (presumably large) boats kept in marinas rather than trailered, there is not much to be said, except that this tendency toward overstatement to some extent counterbalances the understatement in average surplus due to neglect of travel time in the trips model discussed above.

The results for the various trips models appear in table 11.6, where the sample of individuals providing complete responses is 1404. While the overall explanatory power of these models is not particularly impressive, the parameter estimates related to the two proxies for travel coat (MILES or COST) are significant and correctly signed, which is of paramount importance. Secondarily, it appears that larger boating parties, higher educational attainment (a proxy for income) and boat length positively and significantly influence the (logarithm of) the number of boating trips taken per season.

Can we choose between the distance versus coat specifications of the trips models? The models are nonnested, but on the basis of an informal information criterion, mean square error, the decision is almost a toss-up, giving only the slightest edge to the cost specification.[15] While the advantage of this method is computational simplicity, it has no statistical properties. More formal tests, such as those in Pesaran and Deaton (1978) and Aneuryn-Evans and Deaton (1980) are not at all simple. Another alternative is the recent suggestion developed by Davidson and MacKinnon (1981). The logic of this class of tests is developed in appendix A. The application of these tests is also inconclusive, since neither model can be rejected.

In any case, as it turns out the choice of Model I or Model II is not critical for policy evaluation purposes, because both yield similar average surplus measures. Much more important is the choice, given a model

Table 11.6. Alternative Semi-Logrithmic Trips Models (asymptotic standard errors in parentheses)

VARIABLE	Distance Regressor OLS Model II.1.a	Distance Regressor ROBUST[a] Model II.1.b	Cost Regressor OLS Model II.2.a	Cost Regressor ROBUST[a] Model II.2.b
INTERCEPT	2.35676	2.42316	2.35923	2.42832
	(0.13331)	(0.12284)	(0.13330)	(0.12265)
MILES	-0.00082	-0.00097	...	...
	(0.00029)	(0.00026)	...	...
COST	...	...	-0.00823	-0.00983
	...	...	(0.00278)	(0.00251)
PERSONS	0.06277	0.05111	0.06286	0.05107
	(0.01772)	(0.01581)	(0.017711)	(0.01578)
EMPLOY	-0.11228	-0.11819	-0.11242	-0.11869
	(0.07729)	(0.06942)	(0.07727)	(0.06926)
HIGH	0.19821	0.19804	0.19859	0.19799
	(0.07407)	(0.06653)	(0.07405)	(0.06637)
COLL	0.32978	0.32367	0.32974	0.32265
	(0.09723)	(0.08695)	(0.09720)	(0.08675)
PLEASURE	-0.00077	-0.00070	-0.00077	-0.00069
	(0.00082)	(0.00073)	(0.00082)	(0.00073)
HDD	-0.000016	-0.00001)	-0.000017	-0.000012
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
LENGTH	0.01187	0.01103	0.01199	0.01117
	(0.00486)	(0.00470)	(0.00486)	(0.00470)
Mean Square Error	1.2125	0.8772	1.2119	0.8717
Root Mean Square Error	1.1011	0.9366	1.1008	0.9336
R^2	0.031	0.034	0.031	0.034

Notes:

a. Tukey's biweight method estimated by iteratively reweighted least squares. Robust estimates of $\hat{\sigma}$ required to construct weighting function formed from OLS residuals u_i as $\sigma=1.48$ (median$|u_i$ -median $u_i|$) following Holland and Welsch 1977. Respective initial robust estimates for σ are 1.0352 for the distance regressor model and 1.0285 for the cost regressor model.

specification, of OLS or Robust parameter estimates, as evidenced from the average Marshallian consumer surplus estimates given in table 11.7.

There are two interesting observations to be made. First, even with the omission of the opportunity cost of time, the average surpluses per trip per person Prom these trips models appear to be at the high end of the range of similar values reported in the literature. Second, robust regression produces surplus estimates almost 20 percent lower than the OLS estimates. The way outliers are treated has a significant impact on average surplus and ultimately, on the benefits of water quality improvement accruing to the boating category of recreation. So, while the estimates are insensitive to one particular econometric issue, model specification, they are quite sensitive to another, estimation technique.

Table 11.7. Average Surpluses
(average consumer's surplus)[a]

TRIPS MODEL	PER HOUSEHOLD[a]	PER PERSON[b]
II.1.a. DISTANCE, OLS	121.95	38.35
II.1.b. DISTANCE, ROBUST	103.09	32.42
II.2.a. COST, OLS	121.51	38.21
II.2.b. COST, ROBUST	101.73	31.99

Notes:
a. Surplus per household invariant to number of trips taken. See footnote 3.
b. Based on 3.18 persons per trip.

WELFARE ESTIMATES WITH THE RECURSIVE BOATING INTENSITY MODEL

The pollution control benefit component, B_2, accruing to existing boaters can be estimated as the product of the average consumer's surplus per trip (from the semilogrithmic specifications of the preceding section) and the predicted increase in the number of boating trips per boat owning household (from the sequential trips model). All that is needed at this point is a prediction of the change in the number of trips. Heuristically, with the sequential two equation model this involves prediction of the pre and post policy miles per trip using the trailer miles equation and inserting these predicted values into the trips equation to produce predictions of the trips taken pre and post policy.

While in principle the procedure is straightforward, in practice there exist several alternatives for obtaining point estimate predictions when the equations are semilogarithmic. So, analogous to the estimation process itself, judgemental factors can intercede in unforseen ways in the benefit estimation process, even in fairly simple mechanical operations.

Predicting Changes in Miles Travelled and Trips Taken: The Retransformation Problem

The two equations in our recursive system (miles travelled and trips taken) have been estimated with the dependent variables transformed to the logarithmic scale to obtain desirable statistical properties (especially normality) consistent with the assumptions of Cragg's hurdles model. However, pre and post policy predictions are desired on the untransformed scale, and as noted in the literature (Goldberger 1968, Duan 1983, Miller 1984) unbiased and consistent quantities on the transformed scale do not retransform into unbiased or consistent quantities on the untransformed scale.

Suppose the estimated model is:

$$\ln y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{12}$$

where β_0 is an augmented intercept, x_i the independent variable of concern, and ε_i is a normal, independently distributed error term with variance σ^2. It can be shown (Goldberger 1968, Miller 1984) that the exponential retransformation of this model produces a prediction of the conditional median of y_i , not the conditional mean:

$$\text{Med}(y_i) = \exp(\beta_0 + \beta_1 x_i) \tag{13}$$

This conditional median function lies below the conditional mean function by a multiplicative factor which depends on the underlying disturbance variance The conditional mean function is:

$$E(y_i) = \exp(\beta_0 + \beta_1 x_i + \sigma^2/2) \tag{14}$$

If an estimate of the conditional mean is desired, one simple remedy is to replace σ^2 with its sample estimate in Eq. 21, which removes a major portion of the retransformation bias (Miller 1984).

An alternative is to employ Duan's (1983) nonparametric smearing estimate of the bias retransformation factor in lieu of $\exp(\hat{\sigma}^2/2)$. The smearing estimate uses the empirical cumulative density function of the regression residuals $\hat{\varepsilon}_i$ to estimate the required retransformation correction factor as $(\Sigma \exp(\hat{\varepsilon}_i))/n$ where n is the number of observations. The smearing retransformation is:

$$E(y) = \exp(\hat{\beta}_0 + \hat{\beta}_1 X) \times (\Sigma \exp(\hat{\varepsilon}_i))/n \tag{15}$$

The smearing estimate attains high efficiency relative to the parametric normal theory retransformation and provides sane protection against departures form normality.[16]

Thus the predictions of a change in y_i from y_0 to y_1 given a change in x_i Prom pre policy value x_0 to post-policy value x_1 are:

MEDIAN:

$$\text{MED } \Delta y_i = \exp(\hat{\beta} + \hat{\beta}_1 x_0) - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1) \tag{16}$$

MEAN 1:

$$E(\Delta y_i) = \text{NE } (\exp(\hat{\beta}_0 + \hat{\beta}_1 x_0) - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1) \tag{17}$$

where NE represents the naive estimate $\exp\hat{\sigma}^2/2$ and

MEAN 2:

$$E(\Delta y_i) = \text{SE } (\exp(\hat{\beta}_0 + \hat{\beta}_1 x_0) - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1)) \tag{18}$$

where SE represents Duan's smearing estimate $\Sigma \exp(\hat{\varepsilon}_i)/n$.

However, there is a problem with both of the retransformation bias correction factors (NE, SE) which appears to have gone unremarked in the literature. It is that in the presence of the specification error of relevant but omitted explanatory variables (which is frequent in recreational demand modeling due to data deficiencies), the sample estimate of $\hat{\sigma}^2$ is also inconsistent, being above σ^2 in the probability limit (Schmidt 1976).[17] In consequence, the retransformed mean function could well be a biased and inconsistent estimate of the true mean function, exceeding the latter to a greater degree than the median function understates it, depending on the magnitude of the upward bias in $\hat{\sigma}^2$.

To avoid the consequences of an upwardly biased estimate of the conditional mean function it is possible to use a fourth method which employs the ratio of predicted values pre and post policy, thus eliminating the retransformation factor entirely by cancellation.

$$R_i = E(y_1)/E(y_0) = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1)/\exp(\hat{\beta}_0 + \hat{\beta}_1 x_0) \tag{19}$$

Then employing the ratio as a linearization of the estimated function around the originally observed point y_0 to predict y_1, the expected change in y is:

Mean 3:

$$\Delta y = y_0(R_i - 1) \tag{20}$$

Welfare Estimates of the Benefits of Increased Boating Intensity Among Boat Owners Due to Pollution Control

Using a prediction sample of 1174 observations of boat owning households[18] and evaluating the recursive model for each observation pre and post policy produces the average miles response reported in table 11.8, and the average trips response reported in table 11.9. In general the median response theoretically lies below the true mean response which itself should be less than or equal to the smearing and naive retransformation estimates, following the preceding argument. Notably, in tables 11.8 and 11.9 the expected response based on the ratio method, which does not depend on the (potentially biased) sample estimate $\hat{\sigma}^2$ is above the median but below either the smearing or naive retransformation estimates.

As for the effect on round trip travel distance of making all water boatable, it is minimal. From table 11.8 the range of change in expected travel distance due to the policy is a decrease of 1.24 to 2.00 miles, depending on the model estimated and the prediction method (ignoring the median predictions).[19] This small decrease in distance translates into the marginal increases in the number of trips taken per household per year reported in table 11.9.

Since the effect of the policy on quantities (miles, trips) is so slight, it is not surprising that the benefits of the policy emanating from increased intensity of participation among existing boat owners, reported in the last column of table 11.9, are modest. Again ignoring the median, the national benefits range between a low of $20 million per annum to a high of $43 million, demonstrating sensitivity to model and prediction method. The magnitude of this benefit component B_1 relative to the B_2

Table 11.8. Predicted Miles Travelled Per Trip:
The Effect of Pollution Control
(per household)

Miles Travelled Model[a]	Metdod[b]	Miles Pre-Policy	Miles Post Policy	Chance Due to Policy[c]
MODEL I.1.a. STATE DATA, NLIN	Median	29.38	28.83	-0.55
	Mean 1	86.50	84.90	-1.60
	Mean 2	69.23	67.95	-1.28
	Mean 3	72.20	70.95	-1.25
MODEL I.1.b. STATE DATA, ROBUST NLIN				
	Median	31.15	30.57	-0.58
	Mean 1	70.47	69.17	-1.30
	Mean 2	68.11	66.85	-1.26
	Mean 3	72.20	70.95	-1.25
MODEL I.2.a. COUNTY DATA, NLIN				
	Median	37.55	36.92	-0.63
	Mean 1	117.43	115.43	-2.00
	Mean 2	96.03	94.40	-1.63
	Mean 3	72.20	70.96	-1.24
MODEL I.2.b. COUNTY DATA, ROBUST NLIN				
	Median	39.54	38.87	-0.67
	Mean 1	92.65	91.07	-1.58
	Mean 2	94.62	93.01	-1.61
	Mean 3	72.20	70.96	-1.24

Notes:

a. Models from table 11.4.

b. Mean 1 refers to naive retransformation, Eq. 23; Mean 2 to smearing retransformation, Eq. 24; and Mean 3 to ratio method, Eqs. 25-26. True sample mean value of MILES pre-policy is 72.20 for the evaluation sample of 1174 observations, and 68 for the original estimation sample.

c. Equal to POST minus PRE.

Table 11.9. Predicted Boating Trips Per Year: OLS and Robust Tripe Models, Distance Regressor (per household)

TRIPS MODEL	MILES MODEL FOR PREDICT- INC MILES INPUT TO TRIPS MODEL	MILES AND TRIPS PREDIC- TION METHOD[a]	TRIPS PRE- POLICY	TRIPS POST- POLICY	CHANGE DUE TO POLICY[b]	POLICY BENEFIT (million $)[c]
OLS II.1.a.	NONLINEAR:					
	I.1.a. STATE NLIN	Median	14.946	14.953	0.007	6.48-7.68
	I.2.a. COUNTY NLIN	Median	14.860	14.868	0.008	7.41-8.78
	I.1.a. STATE NLIN	Mean 1	26.145	26.178	0.033	30.56-36.20
	I.2.a. COUNTY NLIN	Mean 1	25.670	25.709	0.039	36.12-42.78
	I.1.a. STATE NLIN	Mean 2	25.557	25.583	0.026	24.08-29.18
	I.2.a. COUNTY NLIN	Mean 2	25.126	25.158	0.032	29.64-35.10
	I.1.a. STATE NLIN	Mean 3	24.760	24.782	0.022	20.38-24.13
	I.2.a. COUNTY NLIN	Mean 3	24.760	24.786	0.026	24.08-29.18
ROBUST II.1.b.	NONLINEAR/ROBUST:					
	I.1.b. STATE ROBUST NLIN	Median	15.349	15.358	0.009	8.34-9.87
	I.2.b. COUNTY ROBUST NLIN	Median	15.247	15.256	0.009	8.34-9.87
	I.1.b. STATE ROBUST NLIN	Mean 1	22.910	22.938	0.028	25.93-30.72
	I.2.b. COUNTY ROBUST NLIN	Mean 1	22.564	22.596	0.032	29.64-25.10
	I.1.b. STATE ROBUST NLIN	Mean 2	25.472	25.502	0.030	27.79-32.91
	I.2.b. COUNTY ROBUST NLIN	Mean 2	24.988	25.024	0.036	33.34-39.49
	I.1.b. STATE ROBUST NLIN	Mean 3	24.760	24.782	0.022	20.38-24.13
	I.2.b. COUNTY REBOUST NLIN	Mean 3	24.760	24.786	0.026	24.08-29.18

Notes:

a. Non-robust trailer miles models matched with non-robust trips models, and robust trailer miles models matched with robust trips models in all cases. Same prediction method used for both trailer miles and tripe. Definition as in footnote b, table 11.8.

b. Equal to POST minus PRE.

c. Fraction of boat owners in population (0.1124) time 80 million households times average surplus per trip of either $103 from robust regression (first figure) or $122 from OLS.

component reported in the preceding chapter diminishes the importance or methodological considerations. The benefits of increased intensity of use among individuals owning boats before the quality change appear to be very small.

WELFARE ESTIMATES: BOAT RENTING HOUSEHOLDS

But what about boat renters? This component has thus far been ignored, our analysis focusing instead on ownership for the purely practical reason that the Coast Guard data contains no information on renters other than their numbers, thus prohibiting model estimation. But if renters are assumed to behave just like owners a rough calculation of the rental market component of benefits can be undertaken.

The Coast Guard survey (p. 90) estimates that in 1976 there were 3,752,000 households that rented a boat in 1976, with an average of about 3 trips per household. Using the most generous assumptions, an upper limit on the benefits of water pollution control accruing to new and existing boat renters can be obtained.

The benefits of control accruing to boat renting households, B_R, can be approximated (Vaughan and Russell 1982) by

$$B_R = [OB \times \Delta T + NB \times OT + NB \times \Delta T]V \tag{21}$$

where

OB = the total number of boat renting households before pollution control, 3,752,000

NB = the additional boat renting households attracted by pollution control

OT = the total boating trips per household per year before pollution control, 3

ΔT = the increment to OT from pollution control

V = the consumer's surplus per household per boating occasion.

To get new boating households, NB, suppose the probability of renting a boat increases by the same proportion as our maximum percentage increase in the probability of owning due to pollution control. Using the change in the probability of owning from conditional logit model 17 and a base probability of owning of 0.1124, gives a 1.33 percent increase in the probability of renting. The product of 3,752,000 initial boat renting households and the 0.0133 increment produces 50,000 new boat renting households, NB. Assuming their increase in trips AT is the maximum from table 11.9, 0.039, and a high value for V of $122 per household per trip, we get an upper limit on B_R, in million dollars, of:

$$B_R^{Max} = ((3,752,000 \times 0.039) + (50,000 \times 3) + (50,000 \times .039))122)/(1 \times 10^6)$$
$$B_R^{Max} = \$36.39 \text{ million.}$$

Similarly, a lower limit estimate of R_B can be constructed by assuming only a 0.15 percent increment in the probability of renting (Logit Model 20), or 5,700 new renting households for NB, along with a lower limit of 0.022 for ΔT from table 11.10 and a low value per trip of $103. This gives:

$$B_R^{Min} = ((3,752,000 \times 0.022) + (5,700 \times 3) + (5,700 \times 0.022))103)/(1 \times 10^6)$$
$$B_R^{Min} = \$10.28 \text{ million.}$$

In the next section we combine all benefit components and remark on their magnitude in comparison with previously published estimates.

Table 11.10. Total Boating Benefits of Pollution Control (million 1983 dollars)

	Minimum	Maximum
Increased Ownership[a]	24.62	264.02
Increase Intensity of Use, Owners[b]	20.38	42.78
Increase Rentership[c]	1.76	18.30
Increase Intensity of Use, Renters[c]	8.52	18.09
Grand Total	55.28	343.19

Notes:
a. From table 4.13 chapter 4.
b. From table 11.10, this chapter.
c. Calculated in this chapter.

SUMMARY: TOTAL WATER POLLUTION CONTROL BENEFITS ACCRUING TO THE BOATING CATEGORY OF RECREATION

The analysis of this and preceding chapter has identified four components of boating benefits from bringing all water, both fresh and marine, up to boatable quality. The benefits of increased ownership, increased rentership, increased intensity of use among owners and increased intensity of use among renters have been separately identified and quantified. All of these components are brought together, in summary fashion, in table 11.10, where the minimum and maximum estimate in each category is reported, along with grand totals. Particular combinations of estimates from the model's predictions reported in table 11.13 of the previous chapter, and table 11.9 of this chapter can be chosen, if so desired, to produce point estimates of the grand total lying within the range of $55 to $340 million shown in table 11.10.

The evidence that benefits can vary by a factor of 6 due purely to methodological choices regarding equation specification and estimation procedure, <u>using the same data set</u>, is fairly unsettling. This sort of within-sample uncertainty arising from the selection of a particular set of econometric tools, when coupled with variation across data sets, makes cross-study benefit comparisons a hazardous undertaking. Indeed, most existing studies pay scant attention to within-sample variation in point estimates of benefits, due to the procedural choices open to the investigator, but rather report a single benefit number as if it were fact.[20]

But if comparisons are to be made, reference to Freeman's (1982) most likely boating benefit point estimate of 1.5 billion 1978 dollars per year (with a range of 1 to 2 billion) shows that the benefits reported here are quite small indeed. Moreover, when we recognize that 44 percent of the boaters in the Coast Guard survey reported using their boats more than 50 percent of the time for hunting and fishing rather than just pleasure cruising, the benefits of pollution control accruing to boating per-se are even smaller, the balance being already accounted for in fishing.

It is extremely hard to isolate the primary source of the discrepancy between Freeman's most likely estimate and our own figures. But Freeman's judgement was based on a synthesis of existing studies, most of which appear to have assumed a larger polluiton-reducing policy effect than the 3 percent average improvement reported in the Dyson survey upon which our estimates are based.

NOTES

1. The underlying functions are nonlinear, so evaluation and summation over all individuals is more correct than the aggregate approximation in (4). But in practice the poor quality of the model estimates would not seem to justify function evaluation on an individual-by-individual basis.

2. This procedure is illustrated in Vaughan and Russell 1982, Chapter 6, p. 167. For discussion of the pitfalls, see chapter 3 above.

3. While alternative value per trip estimates could be obtained from published sources, use of such values could be inconsistent with the notion Of a plausibility check based on separate analyses using the same (Coast Guard) data base.

4. For instance, in Vaughan and Russell 1982 a 1979 coat of 7.62 cents per mile was used based on American Automobile Association data, while the U.S. Department of Transportation reports a 1982 cost for intermediate automobiles of 13.5 cents per mile.

5. For an example of this sort of model see Duan, _et_. _al_. 1983.

6. This construct abstracts from the characteristics of the sites visited, which are unknown but may influence the reported distance travelled.

7. The derivative of the function in (11) with respect to b_0 is required:

$$\partial F/\partial b_0 = \frac{(D_F - D_M)(\exp(b_0(D_F/D_M))(D_F/D_M)}{(\exp(b_0(D_F/D_M))D_F + (1-\exp(b_0(D_F/D_M)))D_M}$$

8. With an additive linear form, A can be used directly as a regressor and the unknown factor of proportionality will be absorbed in the parameter estimate attached to it.

9. For estimation using analytical first derivatives of the function in (12) with respect to the parameters.

$$\partial F/\partial b_0 = \frac{(b_1A/D_M)(\exp(b_0b_1A/D_M)(b_1A-D_M))}{(\exp(b_0b_1A/D_M))b_1A + (1-\exp(b_0b_1A/D_M))D_M}$$

$$\partial F/\partial b_1 = \frac{(b_0A/D_M)(\exp(b_0b_1A/D_M))(b_1A-D_M) + (\exp(b_0b_1A/D_M))(A)}{(\exp(b_0b_1A/D_M))b_1A + (1-\exp(b_0b_1A/D_M))D_M}$$

10. Since coat per mile in this model is assumed constant across all individuals, the i^{th} parameter estimates $\hat{\beta}_i$ from an additive model using distance as the i^{th} regressor and a model using a constant, c, times distance, $\hat{\beta}_i^*$ all else constant, are related by (Kmenta 1971, p. 377):

$$\hat{\beta}_i^* = \frac{1}{c}\,\hat{\beta}_i$$

11. It should be noted that many responses were coded 998 or 999 in the data tape which we received without any documentation on coding conventions. Inspection of the raw data revealed these figures, and a call to the Coast Guard confirmed that they represented “no response given” or “don’t know”. Had the raw data not been screened with a preliminary descriptive statistics analysis, these invalid observations could have erroneously been included in the estimation data set.

12. The Tukey biweight robust regression criterion can be implemented using iteratively reweighted least squares (Holland and Welsch 1977). We use the SAS NLIN procedure to minimize:

$$S_{biweight} = \Sigma Q(r),$$

where

$Q(r) = (B^2/2)(1-(1-(r/B^2)^3)$ if $|r|<=B$, or otherwise

$Q(r) = (B^2/2)$

where

B is a tuning constant. We use B=4.685.

r is abs(residual)/$\hat{\sigma}$

$\hat{\sigma}$ is a measure of the scale of the error.

We use the residuals $\hat{u}_i$ from the equal-weights models to construct a robust measure $\hat{\sigma}$=(median$|\hat{u}_i$-median $\hat{u}_i$ $|$)/.6745 which is approximately equal to the true σ if the sample is large and arises from a normal distribution.

The weighting function for the biweight is:

$w_1 = (1-(r/B)^2>^2$ if $|r|<=B$, or

$w_1 = 0$ if $|r|>B$

The biweight estimator depends on both a measure of scale (like the standard deviation) and a tuning constant. Results vary if these values are changed.

Initial starting values for the parameter estimates were obtained from least absolute deviations (LAD) estimation following the weighted OLS procedure suggested by Maddala 1977.

The standard errors reported for the robust regression results are those calculated by the weighted least squares algorithm. Welsch, 1975, suggests that if standard errors are so calculated, the critical tabled value of the “t” statistic employed for hypothesis tests of the robust regression parameters at a particular significance level be adjusted upward by a factor of 1.12. So, for example, the critical value at the 5 percent level becomes 2.20, not the usual 1.96.

13. The expected value nature of the dependent variable, MILES, suggests a possible heteroskedasticity problem, since the standard error of the response may vary inversely with the number of trips taken by an individual. Respondents with a larger sample size (more trips) might report a more precise (lower variance) average distance travelled than those whose average reflects limited trip experience. If this were the case, the standard errors of the parameter estimates reported in table 11.5 would be biased.

Rather than pursuing this suspicion analytically, the residuals from the models in table 11.5 were exposed to the Park (1966) heteroskedasticity test. The test relates variance of the i^{th} observation to trips taken with the general form $\sigma^2 = \sigma^2(\text{TRIPS})^\delta$ where δ is a parameter to be estimated and a consistent estimate of σ_i^2 is constructed as the square of the i^{th} predicted residual. The outcomes of double-log Park test regressions do not reject the null hypothesis of homoskedaaticity, so no remedial procedure is required.

14. No observations received weights leas than 0.40 in the state run and only 5 observations received weights leas than 0.40 in the county run. Moreover, 85 percent of the observations received weights above 0.80 both the state and county runs.

15. This criterion (variously labelled the Sargan teat or Akaike'a Information Criterion) is not really a statistical test with known statistical properties. Instead it is just a method of model discrimination which is easy to calculate and should be successful, according to evidence from Monte Carlo studies, "on average," presuming one Of the models in the comparison set is the true model, No significance level can be set for such a comparison: one just chooses the model with

the higher likelihood (Aneuryn-Evans and Deaton, 1980; Harvey, 1981);

16. If σ^2 is above 2 the relative efficiency of the smearing estimate is low except when, as in our case, the rank of the X matrix is greater than 10. See Duan 1984, table 111.

17. By omitting the travel cost for substitute activities, trips equations like those estimated here obviously could suffer from parameter bias.

18. This sample is smaller than the trips and trailer miles estimation samples because it represents the usable subset with complete data on all variables, including pollution.

19. From the trailer miles expression (12) it is theoretically possible for the derivatives of trailer miles with respect to either distance to freshwater or distance to marine water, ceteris paribus, to take on positive or negative signs. Specifically, if distance to marine water stays fixed while distance to freshwater falls, the expected travel distance can either rise or fall, depending on the initial values of Π_F and Π_M. For some observations in our sample, pollution control has no effect on distance to marine water, so it is conceivable that the unusual result of an increase in trailer miles could result from a decrease in distance to freshwater. However, only four of the observations in our sample exhibited this behavior. Of course, if both distances fall, no problem arises.

20. A previous study of freshwater fishing benefits of pollution control by the authors (Vaughan and Russell 1982) revealed a similar sensitivity to procedure, the benefits varying by a factor of 7 purely due to procedural choices regarding sample design and estimation method (table 6-5, p. 165).

APPENDIX 11.A

TESTING NONNESTED HYPOTHESES USING THE DAVIDSON-MACKINNON TESTS

A family of nonnested hypothesis testing procedures has recently been developed by Davidson and MacKinnon (1981) which are simple to compute -- the C and P tests. With a sufficiently large sample, the idea of this class of tests is to test trips Model I as H_0 against Model II as H_1, conditional on the truth of H_1. Reversing roles, Model II becomes H_0, and is tested against Model I, conditional on the truth of the new H_0.

The set-up for the family of Davidson-MacKinnon tests is quite simple. As before, we have the two competing (nonnested) linear models:

$$H_0\colon \ln Y = f(X,\theta) + \varepsilon_0 \tag{A-1}$$

$$H_1\colon \ln Y = g(Z,\beta) + \varepsilon_1 \tag{A-2}$$

Both error terms are assumed to be normally independently distributed with zero mean and respective variances σ_0^2 and σ_1^2.

Define the maximum likelihood predictions (*) of each observation of the ℓnY vector, given the maximum likelihood (ML) estimates of θ and β, as:

$$f_i^* = f_1(X_i, \theta_{ML}) \tag{A-3}$$

$$g_i^* = g_i(X_i, \beta_{ML}) \tag{A-4}$$

The C (conditional) test of the truth of H_0 involves a linear regression to estimate the test parameter a, conditional on the β_{ML} vector:

$$\ln Y_i = (1-\alpha)f_i^* + \alpha g_i^* + \varepsilon_i \tag{A-5}$$

or

$$\ln Y_i - f_i^* = \alpha(f_i^* - g_i^*) + \varepsilon_i \tag{A-6}$$

The validity of H_0 can be tested by using a conventional t test of the null hypothesis that $\hat{\alpha}$, the estimate of α, equals zero. However, the t statistic for $\hat{\alpha}$ is not distributed asymptotically as N(0, 1) if H_0 is true. Rather, the estimate of the variance of the distribution of the t statistic for the C test is asymptotically biased below 1 when H_0 is true. Practically speaking, this means that the nominal level of significance chosen for the test will overstate the true asymptotic level of significance, or otherwise said, the true probability of Type I error (probability of rejecting a true H_0) will be less than the nominal level chosen. The C test is therefore conservative in the sense that it is less likely to reject a true H_0 than one wishes it to be.

To produce a test statistic which is asymptotically distributed as N(0,1) Davidson and MacKinnon suggest the J (joint) test which estimates a and β jointly:

$$\ln Y_i = (1-\alpha)f_i(X,\beta) + \alpha g_i^* + \varepsilon_i \tag{A-7}$$

However, another test procedure which shares the asymptotic properties of the J test, is the P test. The P test involves a linearization of the J test, around the $\hat{\beta}_{ML}$ vector:

$$\ln Y_i - F_i^* = \alpha(g_i^* - f_i^*) + b_1 f_1' \ldots + b_k f_k' + \varepsilon_i \tag{A-8}$$

where f′ denotes $\partial f/\partial\beta_k|\beta_{kML}$ for k = 1 , ..., K parameters in the model under H_0 and b_1, ..., b_k are parameters to be estimated along with α in the P regression. To complete either the C or P procedures, the roles of H_0 and H_1 are reversed and the tests repeated.

Note that these tests require normally and independently distributed error terms, so before implementation of the tests, it would be prudent to verify the normality assumption, which is also fundamental to the two-step hurdles model employing OLS using the logarithm of the dependent variable in the second (intensity) step.

Thus as a spot check, the OLS residuals from model II.1.a. can be exposed to two normality tests; the nonparametric Kolmogorov test (Conover 1980, Lillefors 1967) and the parametric Kiefer-Salmon score test (Kiefer and Salmon 1983). The relevant test statistics, along with the relevant two-sided critical values at the 5 and 1 percent levels, are given in table 11.A.1. At the 1 percent level, both tests fail to reject normality.

The following outcomes are all possible under the nonnested hypothesis testing scheme.

	Model I	Model II
	$H_0: \ln Y = f(x, \theta) + \varepsilon_0$	$H_0: \ln Y = f(Z, \theta) + \varepsilon_1$
1	Accept	Accept
2	Accept	Reject
3	Reject	Accept
4	Reject	Reject

Only outcomes 2 and 3 provide a judgement between the models.

In this case, the absolute values of the test C and P test statistics are shown in table 11.A.2. On this basis no rejections of either null can be made at the 5 percent level, Thus, these test results are inconclusive. Practically speaking this is not a surprising result, because the difference between the models is only in one variable, COST versus MILES, and these two variables are closely related.

Table 11.A.1. Normality Test Results

Test	Test Statistic	Critical Values[a] 5%	1%
Kolmogorov	0.0331	0.0361	0.0433
Kiefer-Salmon	8.64	5.99	9.21

Note:

a. Kolmogorov critical values from Conover 1980, table A14 for sample size of 1404. Kiefer-Salmon critical values are chi-square with 2 degrees of freedom.

Table 11.A.2. Davidson and MacKinnon Nonnested Test Statistics

	Alternative Hypothesis H_1	
TESTED HYPOTHESIS H_0	Model I	Model II
Model I		
Test C	n.a.	1.41
Test P	n.a.	1.53
Model II		
Test C	1.14	n.a.
Test P	1.26	n.a.

REFERENCES

Allen, R. G. D. 1967. Mathematical Analysis for Economists (New York: St. Martin's Press).

Aneuryn-Evans, G. and A. Deaton. 1980. Testing Linear Versus Logarithmic Regression Models. Review of Economic Studies, 47(146): 275-291.

Bouwes, Nicolaas W. Sr. and Robert Schneider. 1979. "Procedures in Estimating Benefits of Water Quality Change," American Journal of Agricultural Economics, vol. 61, no. 3 (August), pp. 535-539.

Conover, W. J. 1980. Practical Nonparametric Statistics (New York: John Wiley and Sons) 2nd ed. Ch. 6.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica vol. 39 (September) pp. 829-844.

Daniel, Cuthbert, and Fred S. Wood. 1980. Fitting Equations to Data (New York: John Wiley and Sons).

Davidson, Russell and James G. MacKinnon. 1981. "Several Tests for Model Specifications in the Presence of Alternative Hypotheses, Econometrica, vol. 49, no. 3 (May), pp. 781-793.

Duan, Naihua and Willard G. Manning, Jr., Carl N. Morris and Joseph P. Newhouse. 1983. "A Comparison of Alternative Models for the Demand for Medical Care," Journal of Business and Economic Statistics, vol. 1, no. 2 (April), pp. 15-126.

_____. Smearing Estimate: A Nonparametric Retransformation Method". 1983. Journal of the American Statistical Association, vol. 78, no. 383 (September), pp. 605-610.

_____. Willard G. Manning Jr., Carl N. Morris and Joseph P. Newhouse. 1984. "Choosing between the Sample-Selection Model and the Multi-Part Model," Journal of Business and Economic Statistics, vol. 2, no. 3 (July), pp. 283-289.

Dyson, Pamela J. 1984. Recreational Water Availability in the United States: The Impact of Pollution Control (Washington, D.C.: Resources for the Future).

Freeman, A Myrick III. 1982. Air and Water Pollution Control: A Benefit-Cost Assessment. (New York: John Wiley and Sons) 186 pp.

Goldberger, Arthur. 1968. "The Interpretation and Estimation of Cobb-Douglas Functions," Econometrica, vol. 36, no. 3-4 (July), pp. 464-472.

Harvey, A. C. 1981. The Econometric Analysis of Time Series (New York: John Wiley and Sons) Ch. 5.

Hogg, Robert V. 1979. "Statistical Robustness: One View of its Use in Applications Today," American Statistician, vol. 33, no. 3 (August), pp. 108-115.

Holland and Welsch. 1977. "Robust Regression Using Iteratively Reweighted Least-Squares". Communications in Statistics: Theory and Methods, A6(9), pp. 813-827.

Kiefer, Nicholas M. and Mark Salmon. 1983. "Testing for Normality in Econometric Models," Economics Letters, vol. 11, pp. 123-127.

Kmenta, Jan. 1971. Elements of Econometrics (New York: MacMillan).

Lillefors, Hubert W. 1967. "One the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, vol. 62, no. 318 (June), pp. 399-402.

Lin, Tsai-Fen and Peter Schmidt. 1984. "A Test of the Tobit Specification Against an Alternative Suggested by Cragg," Review of Economics and Statistics, vol. 66 (February), pp. 174-177.

Maddala, G. S. 1977. Econometrics (New York: McGraw-Hill).

McConnell, Kenneth E. 1975. "Some Problems in Estimating the Demand for Outdoor Recreation," American Journal of Agricultural Economics vol. 57 (May) pp. 330-334.

McDonald, John F. and Robert A. Moffitt. 1980. "The Uses of Tobit Analysis," Review of Economics and Statistics, vol. 62, no. 2 (May), pp. 328-321.

Miller, Don M. 1984. "Reducing Transformation Bias in Curve Fitting," American Statistician, vol. 58, no. 2 (May), pp. 124-126.

Park, R. E. 1966. "Estimation with Heteroskedastic Error Terms" Econometrica vol. 34, no. 4 (October) p. 888.

Pesaran, M. H. and A. S. Deaton. 1978. "Testing Non-Nested Nonlinear Regression Models", Econometrica 46(3): pp. 677-694.

Schmidt, Peter. 1976. Econometrics (New York: Marcel Dekker).

Schweder, Tore. 1976. "Some 'Optimal' Methods to Detect Structural Shift on Outliers in Regression," Journal of the American Statistical Association, vol. 71, no. 354 (June), pp. 491-501.

Sinden, John A. and Albert C. Worrell. 1979. Unpriced Values: Decisions Without Market Prices. (New York: John Wiley and Sons).

Sorg, Cindy, John Loomis, Dennis M. Donnelly and George Peterson. 1984. "The Net Economic Value of Cold and Warm Water Fishing in Idaho," Unpublished Working Paper, U.S. Forest Service, Collins, Colorado.

Thraen, C.S., J. W. Hammond, and B. M. Buxton. 1978. "Estimating Components of Demand Elasticities from Cross-Sectional Data," American Journal of Agricultural Economics, vol. 60, no. 4 (November), pp. 674-677.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables," Econometrica, vol. 26, pp. 24-36.

U.S. Department of Transportation, U.S. Coast Guard. 1978. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey, Report no. CG-B-003-78, Washington, D.C.

U.S. Department of Transportation. 1982: "Cost, of Owning and Operating Automobiles and Vans," Washington, D.C.: Federal Highway Administration, Office of Highway Planning, Highway Statistics Division.

Vaughan, William J., Clifford S. Russell and Michael Hazilla. 1982. "A Note on the Use of Travel Coat Models with Unequal Zonal Populations: Comment," Land Economics, vol. 58, no. 3 (August), pp. 400-407:

_____ and _____. 1982. "Valuing a Fishing Day: An Application of a Systematic Varying Parameter Model," Land Economics, vol. 58, no. 4 (November), pp. 450-463.

_____ and _____. 1984. "The Role of Recreation Resource Availability Variables in Recreation Participation Analysis," Journal of Environmental Management, vol. 19 (December).

Welsch, Roy E. 1975. "Confidence Regions in Robust Regression," 1975 Proceedings of the American Statistical Association, Statistical Computing Section (Washington, D .C.: American Statistical Association) pp. 36-42.

Wennergren, E. Boyd, Herbert H. Fullerton, John E. Keith, and Robin Meale. 1975. "Economic Value of Water-Oriented Recreation Quality," Utah Water Research Laboratory Study PRRAE 805-1 (Logan, Utah, Utah State University).

Wetzatein, Michael E. and John G. McNeely, Jr. 1980: "Specification Errors and Inference in Recreation Demand Models," American Journal of Agricultural Economics, vol. 62, no. 3 (November), pp. 798-800.

Ziemer, Rod F. and Wesley N. Musser. 1979. "Population-Specific Recreation Demand Models and the Consequences of Pooling Sample Data," Western Journal of Agricultural Economics vol. 4 (July) pp. 121-129.

_____ and _____, Fred C. White and R. Carter Hill. 1982. "Sample Selection Bias in Analysis of Consumer Choice: An Application to Warmwater Fishing Demand," Water Resources Research, vol. 18, no. 2 (April) pp. 215-219.

Chapter 12

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

That there are benefits to society from water pollution control seems intuitively evident. Anyone who has boated on a grossly polluted river or been prevented from swimming or fishing in a nearby lake would agree. Turning this intuitive conviction into monetized estimates, supportable as part of the regulatory process, is an extremely difficult undertaking, however. This report has discussed and critiqued two major methods of doing this: participation equations with separately determined values for days of participation; and demand for goods complementary to water quality, in this case, boats. The results presented have accentuated the critical note. In this final chapter, these results will be brought together and their lessons, if any, for policy and research considered.

SUMMARY OF RESULTS

The results generated in this study are summarized in tables 12.1 and 12.2. In the first of these, methods and data sources are shown, while in the second, benefit estimates are reported. Only the most relentlessly optimistic can take comfort from these numbers. The ranges of uncertainty, viewed simply in dollar terms, are enormous. But more important, those ranges include substantial negative numbers.

As discussed in chapter 9, on the benefits accruing via swimming, boating and mixed swimming and boating participation, some negative benefit estimates are to be expected--or at least need not be disturbing. Thus, participation in boating only might be expected to decline as waters are cleaned up enough to meet the more rigorous requirements of swimming. That is, pollution control might be expected to lead to more mixed-activity days and fewer sole-activity days. But in table 12.2 are found negative benefit estimates covering swimming, boating, and mixed activities together. These results cannot be so easily explained.

The question is, then, what can be made of the results. Three possible answers suggest themselves:

1. There is something wrong with the intuitive expectation that swimming and boating benefits (seen together) must be positive.
2. There is something wrong with the method, either in theoretical structure or estimation technique.
3. The data problems, especially the water quality data problems, described in the text are so severe that the results must be seen primarily as evidence of them.

There does not seem to be any reason to accept the first interpretation. As argued just above, it would not be surprising or disturbing to find negative benefits for some narrowly defined activity category where switching out of the category might be expected to result from water quality improvements. But swimming and boating and mixed-activity days together provide too broad an activity category for this to work.

There are conceptual flaws in the method that involves separate valuation of changes in participation days projected on the basis of socioeconomic variables and aggregated price proxies--whatever the form of the estimated equations. This point is made at considerable length in a companion report (Vaughan et al., 1985). The simulation model developed

Table 12.1. Summary of Method and Data Sources

Activity	Basic method	Key data source	Source of valuation	Chapters
Great Lakes and Marine Recreational Fishing	Participation	NSHFWR '75	Charbonneau and Hay 1978	5, 6, 7
Swimming	Participation	NORS '72	Loomis and Sorg n.d.	8, 9
Recreational Boating	(1) Participation	NORS '72	Loomis and Sorg n.d.	8,9
	(2) Complementary good purchase	U.S. Coast Guard Survey 1973/76	Endogenous travel cost estimates	10, 11

Sources:

Charbonneau, John and Michael J. Hay. 1978. "Determinants and Economic Values of Hunting and Fishing." A paper presented at the 43rd North American Wildlife and Natural Resources Conference, Phoenix, Arizona, March 18-22, 20 pp.

Loomis, John and Cindy Sorg, n.d. "A Critical Summary of Empirical Estimates of the Values of Wildlife, Wilderness and General Recreation Related to National Forest Regions," unpublished, Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station.

U.S. Department of the Interior, Bureau of Outdoor Recreation. 1973. Outdoor Recreation: A Legacy for America (Washington, D.C.), Appendix A.

U.S. Department of the Interior. n.d. 1975 National Survey of Hunting, Fishing and Wildlife-Associated Recreation (NSHFWR) Data Tapes and unpublished mimeo, state and national reports (Washington, D.C.).

U.S. Department of Transportation, United States Coast Guard. 1978. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey, Washington, D.C.

Table 12.2. Summary of Benefit Estimate Ranges by Activity of Method
(10 dollars, 1983, per year)

Activity	Best methods[a] Lowest	Best methods[a] Highest	Simple or naive methods[b] Lowest	Simple or naive methods[b] Highest
Great Lakes and Marine Recreational Fishing	2	425	26	681
Swimming	-313	43	-418	14
Recreational Boating				
(1) Participation	-242	2	-619	-32
(2) Complementary Good Purchase	56	132	136	343
Swimming and Boating				
Joint Participation	-466	432	-910	509
Sum of Boating(2) and Swimming	-257	175	-282	357

[a]"Best Methods" are defined as follows:

- Great Lakes, etc. fishing: Logit form of participation probability, distance form of availability variable, Tukey biweighting in intensity equation.

- Swimming, boating by participation method, and joint swimming/boating: Logit form of participation probability; log form of intensity equation. Distance and density forms of availability variables compared in finding high and low values.

- Boating by complementary good method: County-level water availability; distance and density forms compared in finding high and low values.

[b]"Simple or Naive Methods" are as follows:

- Great Lakes, etc. fishing: OLS form of participation probability, distance form of availability variable, OLS form of intensity equation.

Notes for Table 12.2 (continued)

- Swimming, boating by participation method, and joint swimming/boating: OLS form of participation probability; untransformed intensity variable. Distance and density forms of availability variables compared in finding high and low values.

- Boating by complementary good method: State-level water availability; distance and density forms compared in finding high and low values.

and applied in that report produces results that suggest the possibility of substantial errors in benefit estimates--in either direction and including negative estimates where the true results were positive. Further, it is not possible in the current state of knowledge to predict for real data sets the likely size or direction of errors. Thus, it is not possible here to reject the possibility that the results are simply an artifact of the method.

But it is important to be clear that method is only one half of the benefit scissors. The other half is data. And if available methods leave something to be desired, available data, especially available water quality data, can at best be called inadequate and at worst pathetic. In the absence of comprehensive, consistent, and objectively-based data, it is impossible to test whether the second or third possible explanations is the more appealing. The only even tangentially relevant information is that implied by RFF's earlier study of freshwater recreational fishing. For that study, based on much better water quality data,[1] no problem with negativity was experienced. (A factor of seven difference between high and low estimates <u>was</u> found when variations in methods were explored.)

On this basis, admittedly not an overwhelming one, it nonetheless seems reasonable to conclude that the biggest problem here is with the water quality data. This leads to the major recommendation of the study:

> If water pollution control benefit estimation is to be a serious part of policy making there must be an equally serious effort to build and maintain a water quality database. The two essential characteristics of this database are that it be comprehensive in geographic terms and that the water quality characteristics recorded be relevant. That is, it must be possible to connect the characteristics backward to discharges and forward to activity decisions of individuals.

Such an effort will be important not only to estimation via traditional, indirect methods like the ones exercised here. Even the new and promising contingent valuation techniques require accurate information on actual pre- and hypothesized post-policy conditions in order that respondents can be presumed to be valuing something with consistent, objective meaning.

A second recommendation, however, is that other parts of the data foundation for benefit analysis also be strengthened. In particular, only for recreational fishing is anything approaching a sufficiently large and comprehensive participation data base available. The Coast Guard Boating survey used in chapters 10 and 11 above was gathered with an entirely different purpose in mind and could only be used because a very different method of getting at benefits was adapted from the literature. The NORS data were even less useful, partly because of survey design and partly because of very small sample size. The required data collection effort should encompass at least swimming and boating, should include good information on residence location and on activity venue, and should involve a sample size such that at least 10,000 observations are available for any contemplated estimation effort.

NOTES

1. “Fishable waters” in the base period had been defined and cataloged in a single study; the link between pollution discharges and fishability was carefully, if not completely, defined using biochemical oxygen demand and dissolved oxygen links, and projections of post-policy availability were made on this basis using a reasonable comprehensive model of U.S. pollution sources and major water bodies.

REFERENCES

Charbonneau, John and Michael J . Hay. 1978. "Determinants and Economic Values of Hunting and Fishing." A paper presented at the 43rd North American Wildlife and Natural Resources Conference, Phoenix, Arizona, March 18-22, 20 pp.

Loomis, John and Cindy Sorg, n.d. "A Critical Summary of Empirical Estimates of the Values of Wildlife, Wilderness and General Recreation Related to National Forest Regions," unpublished, Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station.

U.S. Department of the Interior, Bureau of Outdoor Recreation. 1973. Outdoor Recreation: A Legacy for America (Washington, D.C.), Appendix A.

U.S. Department of the Interior. n.d. 1975 National Survey of Hunting, Fishing and Wildlife-Associated Recreation (NSHFWR) Data Tapes and unpublished mimeo, state and national reports (Washington, D.C.).

U.S. Department of Transportation, United States Coast Guard. 1978. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey, Washington, D.C.

Vaughan , William J., John Mullahy, Julie Hewitt, Michael Hazilla and Clifford S. Russell. 1985. "Aggregation Problems in Benefit Estimation: A Simulation Approach," unpublished report to U.S. EPA from Resources for the Future, Washington, D.C. (August).